

# CAN DISTRIBUTED INTERMITTENT RENEWABLE GENERATION REDUCE FUTURE GRID INVESTMENTS? EVIDENCE FROM FRANCE <sup>\*†</sup>

NICOLAS ASTIER<sup>‡</sup>

RAM RAJAGOPAL<sup>§</sup>

FRANK A. WOLAK<sup>¶</sup>

October 11, 2021

## Abstract

This paper estimates the relationship between investments in five distributed generation technologies and hourly net injections to the distribution grid for over 2,000 substations in France between 2005 and 2018. We find that investments in distributed wind and solar capacity have little or no impact on the annual peak of hourly net injections to the distribution grid, while investments in hydroelectric and thermal distributed generation significantly reduce it. An optimistic analysis of battery storage suggests that high levels of investments are required for distributed wind and solar investments to deliver similar reductions in the annual peak of hourly net injections.

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 and Arizona State University for their feedback and suggestions. Remaining errors are ours.

‡Precourt Institute for Energy (Bits & Watts initiative), Stanford University, Stanford, CA 94305, nicolas.astier@stanford.edu

§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. For the most part, whether these wind and solar resources should be installed in the distribution network or in the transmission network depends on the extent to which distribution network connected generation units reduce the need for future transmission and distribution network investments. Large-scale wind and solar generation units connected to the transmission grid produce electricity at a significantly lower levelized cost than wind and solar units connected to the distribution network (IRENA, 2020).<sup>1</sup> Consequently, unless distribution-connected renewable generation units yield significant savings in future transmission and distribution (T&D) network investments, large-scale wind and solar generation units connected to the transmission network is the least cost source of renewable electricity.

The extent to which investments in distributed generation, particularly rooftop solar photovoltaic (PV), reduce the need for future T&D network investments is however highly debated. Many consultant reports claim significant avoided costs of future network investments associated with deploying distributed solar PV capacity.<sup>2</sup> Because T&D network costs typically comprise at least one-third of elec-

---

<sup>1</sup>For example, rooftop solar systems can be up to five times more costly than utility-scale systems on a per unit of generated electricity basis (<https://www.lazard.com/perspective/levelized-cost-of-energy-and-levelized-cost-of-storage-2020/>).

<sup>2</sup>Muro and Saha (2016) survey several such studies.

tricity bills, the claimed savings could easily add up to billions of dollars annually for a country like the U.S.<sup>3</sup> Yet, engineering studies often reach less optimistic conclusions (Cohen and Callaway, 2016; Cohen et al., 2016). Despite this large uncertainty regarding the network deferral benefits from distributed resources, public policies typically strongly support distributed generation (e.g. through higher per-MW subsidies and/or net metering tariffs). As a result, very substantial investments in these technologies have occurred over the past decade. Between 2010 and 2019, global investments in distributed and utility-scale solar were indeed comparable, averaging \$59 billion annually for each type of installations (IRENA (2020), Figure 1.12). For example, about a third of solar electricity generation in the United States today is coming from distributed units.<sup>4</sup>

We contribute to this debate by taking advantage of a unique panel dataset of hourly net injections to over 2,000 local distribution networks throughout France for fourteen years during which over 25 gigawatts (GW) of distributed generation investments occurred. We use this dataset to estimate how different quantiles of the annual distribution net injections at a distribution network substation are impacted by investments in five types of distribution generation technologies. Specifically, we assess the impact of a 1 MW increase in each of five distributed generation

---

<sup>3</sup>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](http://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).

<sup>4</sup>[https://www.eia.gov/totalenergy/data/monthly/pdf/sec10\\_10.pdf](https://www.eia.gov/totalenergy/data/monthly/pdf/sec10_10.pdf).

technologies – solar PV, wind, small hydro, renewable thermal, and non-renewable thermal – on the percentiles of the annual distribution of hourly net withdrawals from the distribution network, and on the percentiles of the annual distribution of hourly ramp rates, defined as differences between two consecutive hourly net injections to the distribution grid.

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

A 1 MW investment in distributed solar PV or wind capacity predicts similar absolute value changes in both extremes of the annual distribution of the hourly ramp rates of net injections to the distribution grid. A 1 MW increase in either wind or solar PV distributed generation capacity predicts a 0.15 MW<sup>5</sup> decrease in the 1st percentile of the annual distribution of the hourly ramp rates. For the 99th percentile, a 1 MW increase in wind or solar PV capacity predicts a 0.14 MW increase. For each of the remaining three distributed generation technologies, a

---

<sup>5</sup>We compute hourly ramp rates as the difference between two consecutive hourly net load levels. Because we measure changes in load levels (measured in MWh) over the course of 1 hour, we use MW as the unit for hourly ramps.

1 MW increase in installed capacity does not predict a non-zero change in any percentile of the annual distribution of hourly ramp rates.

We then test for the possibility of decreasing marginal benefits from distributed generation investments. We find that an additional 1 MW of distributed wind capacity reduces the highest percentiles of the annual distribution of hourly net withdrawals more at lower levels of penetration of wind capacity. The first MW of distributed wind capacity installed at a substation is indeed predicted to reduce the 99th percentile of the annual distribution of hourly net injections to the distribution grid by 0.07 MWh, 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 not statistically different from zero even at low levels of penetration of solar capacity. Distributed solar is however found to reduce the lowest percentiles of the annual distribution of hourly net withdrawals significantly more at higher levels of penetration.

Finally, we explore whether battery storage investments bundled with wind and solar distributed generation could deliver annual peak net withdrawal reductions similar to those obtained for hydroelectric and thermal distributed generation investments. We show that, for this to be the case, storage investments substantially higher than existing levels would have to accompany investments in distributed solar and wind generation. Specifically, we find that in order to achieve reductions in the 99th percentile of the annual distribution of hourly net withdrawals close to

0.1 MWh per MW of distributed wind and solar generation capacity, one Tesla Powerwall 2 battery would need to be installed for every 3 kilowatts of distributed wind or solar capacity installed in the distribution grid. Because 3 kilowatts is the smallest rooftop solar system typically installed, this means that virtually every rooftop solar system would have to be accompanied by a Tesla Powerwall 2 battery.

Taken together, our results argue that, at least for the case of France, increases in distributed solar and wind capacity are more likely to require increases, rather than decreases, in future network investments, unless they are accompanied by very substantial investments in distribution network connected storage devices. Investments in distributed wind and solar without any storage capacity are indeed predicted to result in small or zero reductions in the highest percentiles of the distribution of hourly injections to distribution grids, which drive how much grid capacity is ultimately installed. In addition, distribution network investments may be necessary to manage the increase in the extremes values of the annual distribution of hourly ramp rates for net injections to the distribution grid associated with investments in distributed wind and solar generation capacity. Our results imply that the benefits from deferring future T&D expansions cannot rationalize a substantial policy support for distributed wind and solar generation over utility-scale generation.

The remainder of the paper is organized as follows. The next section discusses the role of distributed generation in the energy transition. Section 3 describes why investments in distributed generation units may reduce the need for future invest-

ments 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**

Perhaps the best way to illustrate the impact of distributed generation on the electricity supply industry is that a Google search of the keywords “distributed generation” yields about half of a billion search results. The tremendous interest in distributed generation is somewhat surprising because the levelized cost of energy from a distributed wind or solar generation unit is higher than a grid-scale wind or solar generation unit (IRENA, 2020) and both types of generation resources produce zero carbon electricity. Therefore, distributed generation units must provide additional benefits in order to be cost-competitive relative to grid-scale facilities (Burger et al., 2019).

The two major benefits claimed are greater customer control over their electricity supply and 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 load, where net load is defined as gross consumption minus distributed generation.<sup>6</sup>

In the early 2000s, distributed generation mostly consisted in dispatchable thermal units. Studies assessing the T&D deferral potential of distributed generation hence assumed the capacity of distributed generation units to be firm, that is to be available whenever it may be 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. The external validity of the simulations performed to actual electricity supply industries is particularly difficult to assess. Indeed, these studies typically focus on a single or a handful of distribution systems, and optimize the installation and sizing of distributed generation instead of observing actual investments.

These limitations point to the need for empirical evidence based on large-scale observational data, as we provide in this work. Borenstein (2008) argues qualitatively that the T&D benefits from rooftop PV should be small but notes the lack of academic literature on the topic.<sup>7</sup> Even the most recent studies that explore the spatial heterogeneity in the social value of intermittent renewable technologies

---

<sup>6</sup>“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).

<sup>7</sup>Similarly, Sexton et al. (2021) acknowledge in their footnote #5: that “evidence on distribution cost avoidance by distributed solar capacity is scarce”.

(Callaway et al., 2018; Fell et al., 2021; Sexton et al., 2021) do not account for the sub-transmission and distribution grids, notably because of the difficulty in gaining access to the necessary data for a large service territory.

We were able to identify only two studies that 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 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, this study is the first to rely on a dataset rich enough to reach a sufficient level of statistical power despite using exclusively observational data and a conservative inference strategy. Indeed, because distribution systems are usually operated by local monopolies, the geographical scope of available data tends to be limited in U.S. case studies (e.g. PG&E in Cohen et al. (2016) or a local utility in Ovaere et al. (2020)). In contrast, a single distribution system operator (DSO) supplies 95% of end-consumers in France, which allows access to internally consistent data on the use of distribution systems across a very large service territory. Finally, observing the hourly utilization of distribution grids enables us to implement a novel empirical strategy that looks at the full distribution of hourly power flows, to study hourly

ramps and to simulate the impact of rolling-out battery storage.

### **3 Distributed Generation and the Electricity Grid**

Electricity supply industries still largely consist of generating electricity at large-scale facilities that take advantage of economies of scale in production, and then transmitting it at a 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. 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 grid power flows.<sup>8</sup> Distributed generation capacity 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.

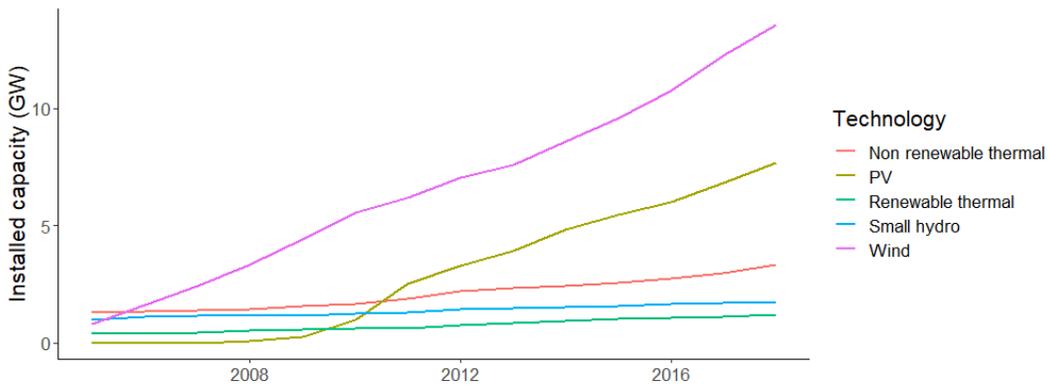
Distributed generation installations generally belong to one of the following categories: wind, solar, small hydro, and thermal units using either renewable (e.g. wood, waste) or non-renewable (e.g. natural gas, diesel) fuels. Figure 1 shows

---

<sup>8</sup>With sufficient distributed generation capacity, electricity may even no longer always flow from large-scale generation units to final consumers during some hours.

how the installed distributed generation capacities that we observe for France have evolved between 2005 and 2018.<sup>9</sup> Although all technologies exhibit an upward trend, solar and wind have by far experienced the most significant growth. As of 2018, there was approximately 28 GW of distributed generation in France.<sup>10</sup>

Figure 1: Total installed capacities of distributed generation (in GW) by year and technology in France, as observed in our final dataset (see Section 4).



Investments in distributed generation units in France have benefited from different types of government support over our sample period.<sup>11</sup> These support mechanisms 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 for dis-

<sup>9</sup>As discussed in Section 4, our final dataset includes the vast majority of distributed generation units located in mainland France.

<sup>10</sup>As a comparison, the highest demand for grid-supplied electricity ever recorded in the country is about 100 GW.

<sup>11</sup>For more details, see for example [www.cre.fr/Transition-energetique-et-innovation-technologique/soutien-a-la-production/dispositifs-de-soutien-aux-enr](http://www.cre.fr/Transition-energetique-et-innovation-technologique/soutien-a-la-production/dispositifs-de-soutien-aux-enr).

tributed generation investments. Land and resource availability considerations can create incentives to locate distributed generation units non-uniformly across space, which we will control for through distribution network fixed effects. We will also control for national trends such as load growth and differences over time in distributed generation support mechanisms with year-of-sample fixed effects.

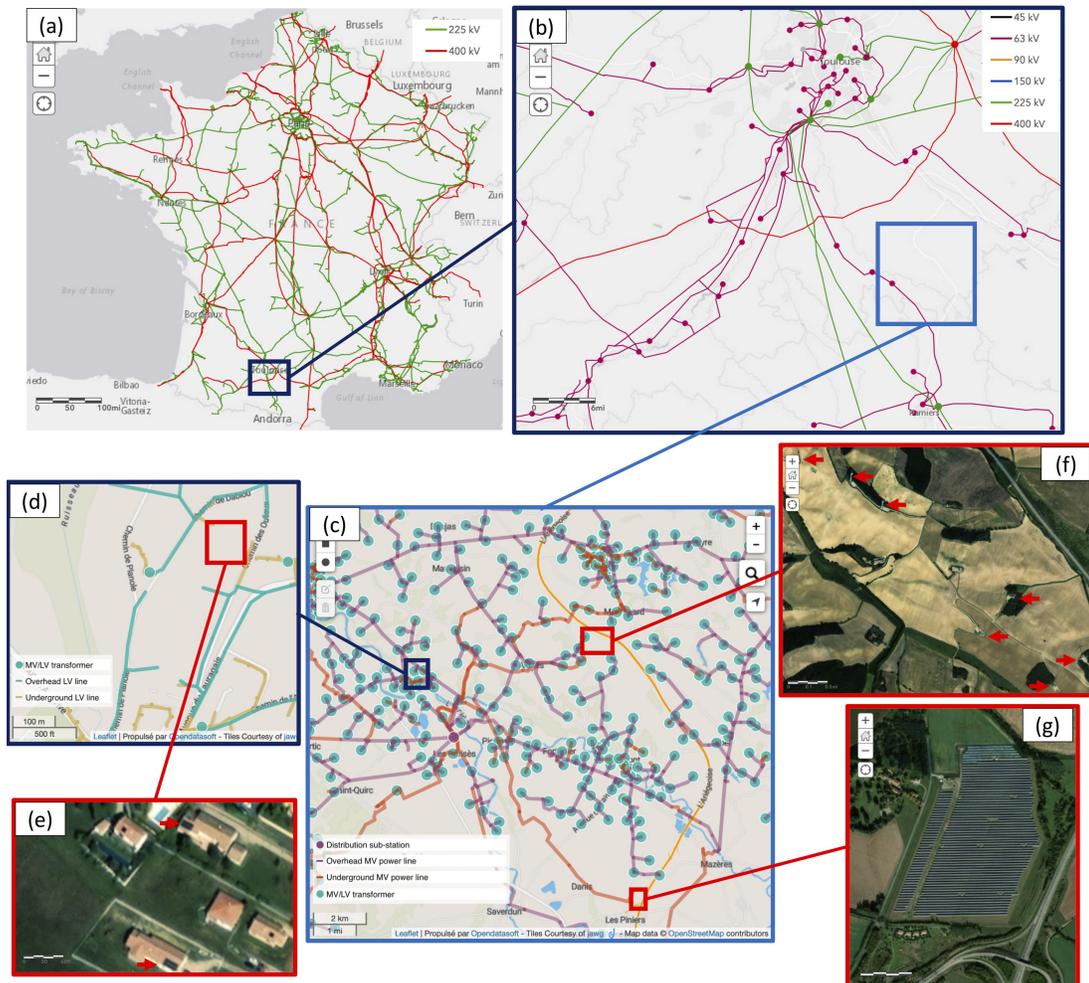
Figure 2 illustrates the overall structure of the French electricity grid,<sup>12</sup> and highlights examples of distributed generation units. Importantly, these units connect to the electricity grid either “behind-the-meter” on the customer’s premises (panel (e)), or through a medium-voltage power line in the distribution network (panels (f) and (g)). A direct approach to study the impact of distributed generation investments on future grid investments is to look at power flows at the interface between the transmission and distribution grids. This interface consists of assets called distribution substations, represented by the purple dots on panel (b) of Figure 2. In France, these substations typically host 63/20kV or 90/20kV transformers.<sup>13</sup> Although the exact voltage levels used in the electricity industry differ across countries, distribution 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). Because the power flows measured at a given substation aggregate all the consumption and generation of users that connect to the grid downstream from the substation, we

---

<sup>12</sup>Panels (a), (b), (e), (f) and (g) are screenshots taken from: [www.arcgis.com/home/webmap/viewer.html?webmap=02d413bcbe184384ba424fc40f9b8ce8](http://www.arcgis.com/home/webmap/viewer.html?webmap=02d413bcbe184384ba424fc40f9b8ce8). Panels (c) and (d) are screenshots taken from: [www.enedis.fr/cartographie-des-reseaux-denedis](http://www.enedis.fr/cartographie-des-reseaux-denedis)

<sup>13</sup>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).

Figure 2: (a) French transmission grid (225 kV and 400 kV power lines) as of 1 September 2016 ; (b) Zoom on the high voltage (63 kV) sub-transmission grid for a given area ; (c) Zoom on the area around a single distribution substation (purple dot), with the corresponding medium voltage (20 kV) power lines and MV/LV transformers ; (d) Zoom on a given neighborhood with its MV/LV transformers and LV power lines. Panels (e)-(f) show examples of distributed generation units connected to the substation. Panel (e) highlights two houses with residential PV systems (about 5-6 kW each). Panel (f) is a 14.3 MW wind unit (arrows point to the 7 wind generators). Panel (g) is a 10.5 MW utility-scale PV unit (sources: ArcGIS, RTE, Enedis).



refer to these power flows as *net* load levels, with the convention that positive values correspond to hours where local consumption exceeds local generation.

To understand how distributed generation affects the future need for network capacity, we build on the power systems literature. More specifically, we study how the percentiles of the annual distribution of the hourly net load levels at each substation change in response to investments in different distributed generation technologies. The percentiles of the annual distribution of the hourly net load levels indeed map to what is known in the power systems literature as the load duration curve. Up to a change in the direction of the x-axis, the load duration curve corresponds to the inverse cumulative distribution function of hourly net load levels. This curve is used to assess the probability that net load may exceed a given level. Planning rules then set a probability threshold  $\hat{p}$  to decide on the size of a given grid component. Whether or not distributed generation decreases the grid capacity needed to ensure a given level of reliability of supply depends on the extent to which this electricity is produced during local peak hours. Figure 3 illustrates this intuition. If the output from distributed generation is not coincident with local peak demand hours, 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).

In recent years, power system engineers have also paid increasing attention to high variations in net load levels over short periods of time, known as “ramps”.<sup>14</sup>

---

<sup>14</sup>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

Figure 3: 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.

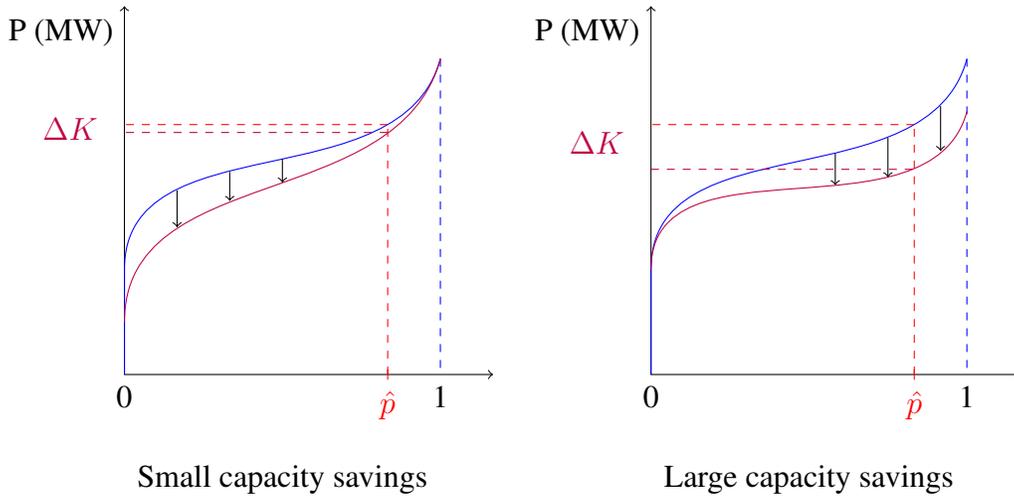
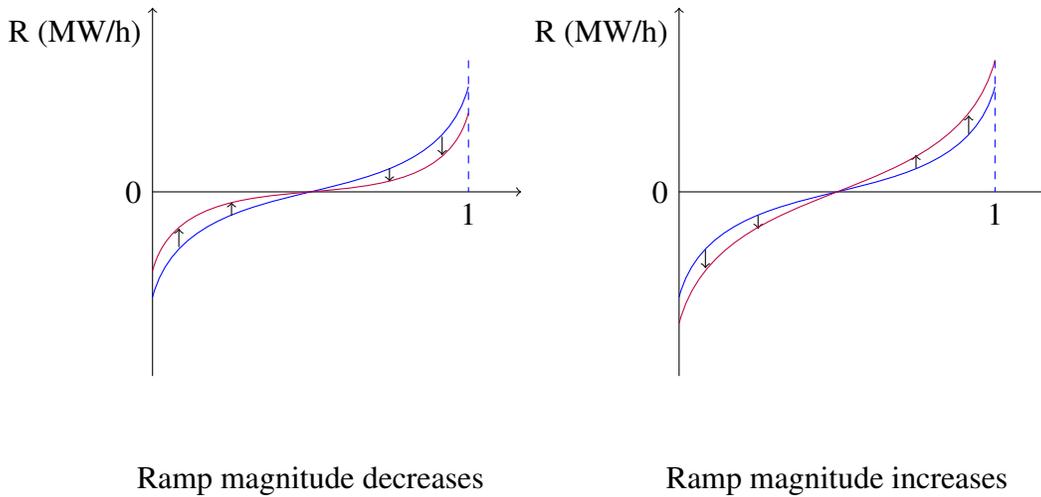


Figure 4: Illustration of how distributed generation may shift the ramp duration curve of a substation, decreasing or increasing the magnitude of hourly ramps.



We thus also look at the impact of distributed generation on the annual distribution of hourly ramps. For a given substation in a given year, the hourly ramp in hour  $h$  ([www.caiso.com/documents/flexibleresourceshelprenewables\\_fastfacts.pdf](http://www.caiso.com/documents/flexibleresourceshelprenewables_fastfacts.pdf)).

is defined as the difference between the load level in hour  $h + 1$  and the load level in hour  $h$ . Hourly ramps can then be sorted in increasing order to build a ramp duration curve, and one can investigate how this curve changes with increased levels of distributed generation. By construction, the integral of the ramp duration curve is close to zero.<sup>15</sup> As a result, the ramp duration curve starts at negative values, which correspond to hours during which net load is decreasing at the highest rates, and ends at positive values, which correspond to hours during which net load is increasing at the highest rates. The flatter the ramp duration curve, the less severe are the observed ramps. In particular, the two extremities of the ramp duration curve materialize the largest variations in net load levels (in absolute value terms). Figure 4 illustrates two contrasted ways in which distributed generation may impact the ramp duration curve faced by a given substation. In both panels, the blue curve represents the pre-existing ramp duration curve and the purple curve the ramp duration curve after the addition of 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). In the latter case, operating the distribution system 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

---

<sup>15</sup>If  $R(h) \equiv L(h + 1) - L(h)$  is the hourly ramp for hour  $h$  (where  $L(h)$  is the net load level in hour  $h$ ), then  $\sum_{h=1}^{8759} R(h) = L(8760) - L(1)$ . This difference is negligible relative to the total load  $\sum_{h=1}^{8760} L(h)$  supplied by the substation over the course of the year.

balancing).<sup>16</sup>

## 4 Data

### 4.1 Substation Hourly Net Load Levels

We observe hourly net load levels for 2,216 substations<sup>17</sup> in France (shown in Figure 5) between 1 January 2005 and 31 December 2018. From this raw dataset of 250+ million observations, we can compute summary statistics for the distribution of hourly net load levels at a given substation in a given year. Appendix C provides detailed information on a number of these summary statistics. In particular, we observe that the most prominent changes that occurred between 2005 and 2018 relate to reverse power flows, that is to hours during which local generation exceeds local consumption. 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. 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

---

<sup>16</sup>For more details on power systems operations, see for example Kirschen and Strbac (2018).

<sup>17</sup>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.

2005 to almost 9% in 2018.

Figure 5: Location of the distribution substations for which we observe hourly net load levels.



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 load levels. 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 load levels (resp. hourly ramps). Our data on observed power flows thus consist of 14 panels (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 information on the universe of power plants in France.<sup>18</sup> 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,<sup>19</sup> distributed generation units belong to one of the five following categories: wind, solar,<sup>20</sup> small hydro, and thermal units burning either renewable (e.g. biomass) or non-renewable (e.g. gas) fuel. We also observe the installed capacities<sup>21</sup> and commissioning dates<sup>22</sup> of distributed generation units.

---

<sup>18</sup>[www.data.gouv.fr/en/datasets/registre-national-des-installations-de-production-deelectricite-et-de-stockage-au-31-decembre-2018/](http://www.data.gouv.fr/en/datasets/registre-national-des-installations-de-production-deelectricite-et-de-stockage-au-31-decembre-2018/). The dataset was downloaded on 28 August 2020.

<sup>19</sup>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.).

<sup>20</sup>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”.

<sup>21</sup>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.

<sup>22</sup>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

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).

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 capacity.

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.<sup>23</sup> 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.

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), broken down by technology. Because the installed capacities of distributed generation were small in 2005 (Figure 1), many observations are at zero. Nevertheless, we observe significant variation in the installed capacity of distributed generation at the substation level in each year, as shown in Table 2. The last

<sup>23</sup>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<sup>2</sup> (4.6 miles squared).

column in Table 2 shows the total installed capacity by technology in our dataset as of 31 December 2018, both in absolute value and as a fraction of the total distributed generation capacity listed in the public inventory of power plants. We note that the substations for which we observe hourly net load levels are hosting the vast majority of distributed generation installations in mainland France.<sup>24</sup>

## 5 Empirical Strategy

We characterize changes in the load duration curve (resp. ramp duration curve) induced by investments in a given distributed generation technology through a quantile impact function. This function captures the average impact that adding 1 MW of capacity has had on individual quantiles of the annual distribution of hourly net load levels (resp. hourly ramp rates) at the substation level. For a given technology, the quantile impact function maps each quantile index (from 0 to 1) to the average effect (in MWh or MW) that adding 1 MW of this technology has had on the corresponding quantile of the annual distribution of hourly net load levels (resp. hourly ramp rates, see Figure 6).

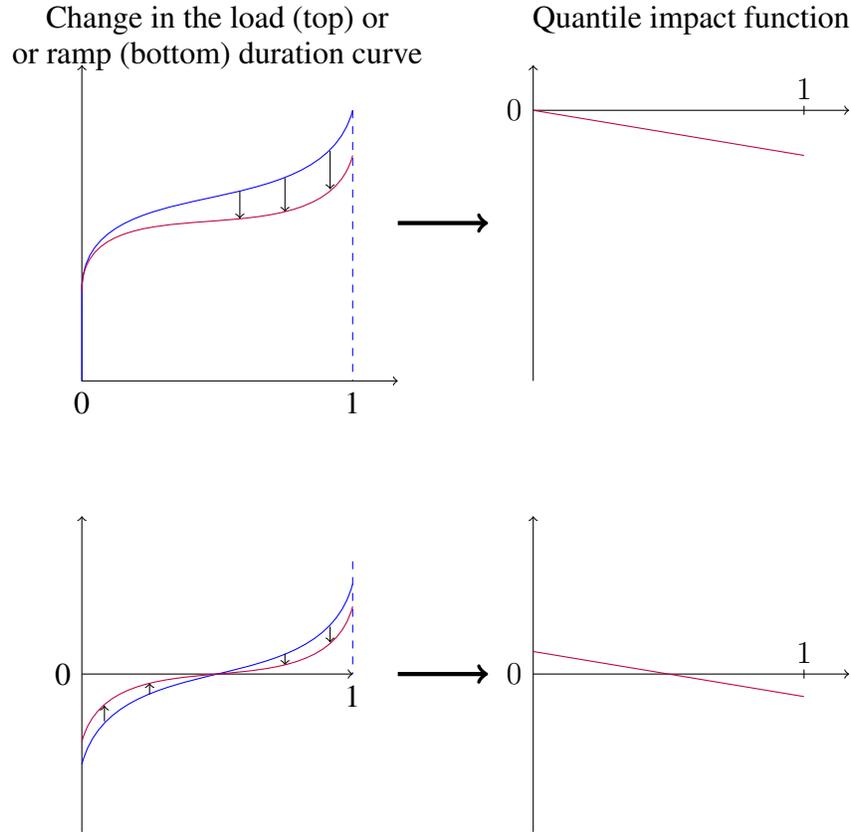
We use a seemingly unrelated regressions framework with a two-way fixed effect model to estimate quantile impact functions.<sup>25</sup> In other words, we run the follow-

---

<sup>24</sup>The remaining capacities most likely connect to substations for which we do not observe hourly net load levels. For example, a number of small hydro units are located on the Northern part of the Rhine river, which is one of the few areas where we lack information on substation hourly net load levels (see Northeastern region on Figure 5).

<sup>25</sup>We discuss in Appendix B the results obtained with alternative specifications. In particular,

Figure 6: Illustration of the intuition behind quantile impact functions for the load duration curve (top) and the ramp duration curve (bottom).



ing 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 load levels on the one hand, and the annual distribution of hourly ramps on the other hand:

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

where  $Y_{q,s,y}$  denotes the  $q$ -th quantile of either the annual distribution of hourly

adding interaction terms or region-year fixed effects were not found to change our results, justifying our use of a more parsimonious model. Non-constant marginal impacts are discussed in Section 7.

net load levels or the annual distribution of hourly ramps for substation  $s$  in year  $y$ ,  $K_{t,s,y}$  the installed capacity of distributed generation technology  $t$  connected to substation  $s$  as of year  $y$ , and  $\delta_s$  and  $\delta_y$  are respectively substation and year fixed effects. We thus estimate fourteen linear regressions (7 percentiles for the load duration curve and 7 percentiles for the ramp duration curve) using ordinary least squares. 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 MWh or MW) 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 load level supplied by distribution substations by 0.2 MWh.

## 6 Main Results

We first estimate the impact of the different distributed generation technologies on the quantiles of the annual distribution of net load levels at the substation level. Table 3 shows our estimation results, which are also represented graphically on Figure 7. 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.

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

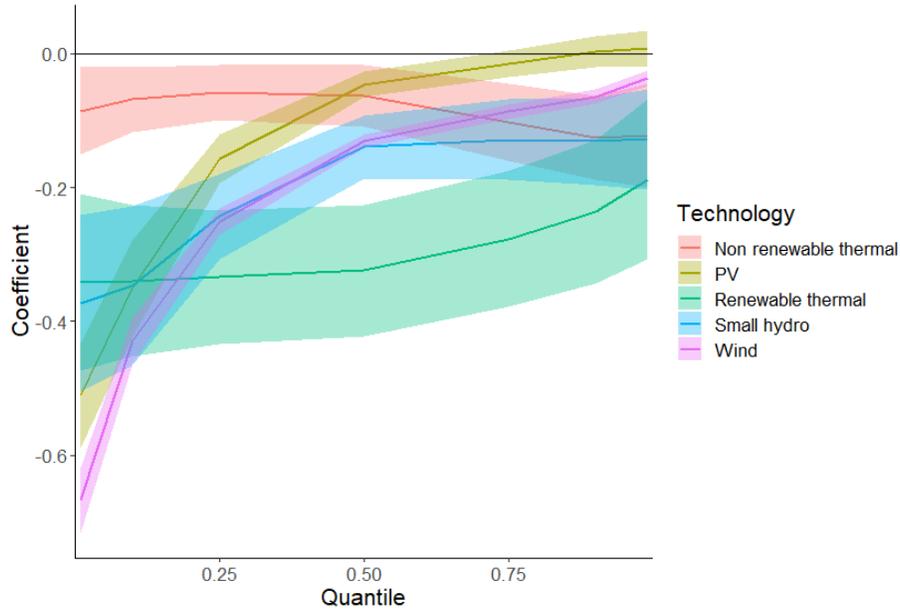
Table 3: Estimated coefficients when regressing the main quantiles of the distribution of hourly net load levels (for a given substation in a given year) on the installed capacities of the different technologies. Robust standard errors clustered at the substation level are reported.

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

more a given technology is associated with a decrease in the peak net load supplied by substations, and thus the more likely it is to defer or avoid grid expansions. In France, the national annual peak load 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 load levels. The impact of wind on the top quantiles is also very small even though, as reported in RTE (2018), average wind generation is higher during the winter. The other technologies are found to have a sizable impact on local peak net load, with an average impact on the 99th quantile of at least 0.12 MWh per MW of distributed generation.<sup>26</sup> Conversely, large

<sup>26</sup>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/](http://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033385467/)).

Figure 7: Graphical representation of the quantile impact functions for the distribution of hourly net loads. Thick lines correspond to the point estimates. Sleeves delimit (two-sided) 5% confidence intervals from robust standard errors clustered at the substation level.



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 load levels.

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 load levels supplied by the substation, increasing the utilization

rates of 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 load levels, that is to expand the range of net load levels that must be supplied by the substation. 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 7 suggests that quantile impact functions are monotone increasing for all technologies but non-renewable thermal. We 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 non renewable thermal, and the null hypothesis of a decreasing quantile impact function cannot be rejected at the same level of statistical significance.

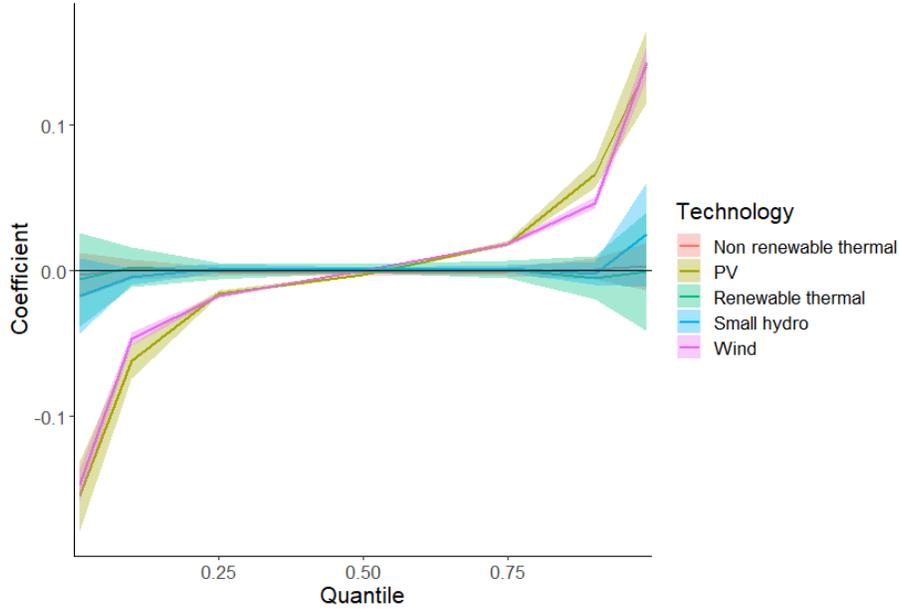
Next, we estimate the impact of the different distributed generation technologies on the ramp duration curve faced by substations. Table 4 shows our results, which are represented graphically in Figure 8. Two distinct groups of distributed generation technologies emerge. On the one hand, thermal and 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 capacity of either wind or PV is associated with an average increase of 0.14-0.15 MW 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 load levels at the substation level. Such an increase 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 DSOs 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.

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.

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

Figure 8: Graphical representation of the quantile impact functions for the distribution of hourly ramps. Thick lines correspond to the point estimates. Sleeves delimit (two-sided) 5% confidence intervals from robust standard errors clustered at the substation level.



## 7 Discussion

### 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 paragraph, we explore whether the marginal impacts of distributed wind and solar on the duration curves change with the level of installed generation capacities. More

specifically, we estimate the following quadratic specification:

$$Y_{q,s,y} = \sum_t \alpha_{q,t} K_{t,s,y} + \sum_t \beta_{q,t} K_{t,s,y}^2 + \delta_s + \delta_y + \epsilon_{s,y} \quad (2)$$

where  $Y_{q,s,y}$  is quantile  $q$  of the distribution of hourly net load levels/ramps for substation  $s$  in year  $y$ . Tables E.8 and E.9 in Appendix E report the obtained 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 load levels is constant 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 distributed solar generation is increasingly responsible for the lowest hourly net load levels 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 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 accelerating rate.

We also explored the potential interaction effects of investments in the different distributed generation technologies in Appendix B. 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.

## 7.2 Battery Storage

This paragraph explores the extent to which battery storage may enhance the grid deferral benefits from distributed wind and solar generation. A given storage system is characterized by two main metrics: the maximum amount of energy  $E$  it is able to store (in kWh), and its maximum charge/discharge rate  $P$  (in kW).<sup>27</sup> The charge rate or “C-rate” (in  $\text{h}^{-1}$ ) 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.

To assess the extent to which battery storage may increase the capacity deferral benefits from distributed wind and solar generation, we assume that for each kW

---

<sup>27</sup>We assume for simplicity that the maximum charge and discharge rates are equal although this need not be the case in practice.

of distributed wind or solar generation connected to a substation in a given year, a proportional amount of storage is also installed. Because (battery) storage “capacity” is generally understood as the maximum energy that can be stored (rather than the maximum (dis)charge rate), we define “storage penetration” as the ratio of installed storage (in kWh) and installed distributed generation (in kW).<sup>28</sup> 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.<sup>29</sup> Battery storage sales in California in 2020 were expected to be 50,000 systems<sup>30</sup> along with expected investments in about 1.1 GW of residential rooftop PV for that year.<sup>31</sup> This suggests that the current battery storage penetration for new residential PV installations is 60% assuming each new battery storage system is a Tesla Powerwall 2.<sup>32</sup>

---

<sup>28</sup>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. The corresponding level of storage penetration is obtained by multiplying the level of penetration we consider here by the C-rate. For example, for a penetration level of 100% (in energy) of a battery storage technology with a C-rate of 0.5, each kW of wind/PV capacity is bundled with 1 kWh of storage, which adds 0.5 kW to the maximum (dis)charge rate. This situation hence corresponds to a penetration level (in terms of maximum (dis)charge rate) of 50% ( $= 0.5 \times 100\%$ ).

<sup>29</sup>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.

<sup>30</sup><https://about.bnef.com/blog/california-household-battery-sales-to-quadruple-in-2020/>

<sup>31</sup>[www.seia.org/sites/default/files/2021-03/California.pdf](http://www.seia.org/sites/default/files/2021-03/California.pdf)

<sup>32</sup>Note that  $60\% \approx (50,000 \text{ batteries} \times 13.5 \text{ kWh/battery}) / (1,100,000 \text{ KW of new solar PV})$ .

Finally, we assume that batteries are operated in a weekly peak-shaving mode.<sup>33</sup> In other words, the storage operator aims at minimizing either weekly peak net demand 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. 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.

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 three additional conservative assumptions. First, we neglect for simplicity power losses when charging/discharging electricity. Second, we assume that the storage operator has a perfect foresight of substation hourly net load for the upcoming week. Third, we do not model battery degradation and the opportunity cost of being able to enjoy a limited number of refresh cycles. Overall, our aim is to get a sense of the upper bound of the potential grid deferral benefits from combining distributed wind

---

<sup>33</sup>We also ran the same analysis with storage operated in a daily peak-shaving mode, which mechanically yields lower reductions in peak hourly net load levels. The impact of PV on the 99th percentile was found to reach -0.05 MWh per MW at 500% penetration of battery storage. The impact of wind on the 99th percentile was found to increase to -0.05 MWh per MW at 50% penetration of storage, but not to increase further at higher levels of storage penetration. This suggests that substation peak net load is typically reached during days with consistently low wind output.

and solar with battery storage. Operating batteries in a peak-shaving mode seems likely to be the most critical assumption.<sup>34</sup> The perfect forecast assumption is also quite optimistic, but building a credible hourly net load forecast model is 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% range, and peak-shaving would in practice be needed during a limited number of hours to shave only the yearly peak in hourly net load levels.

Figure 9: Obtained quantile impact functions for the distribution of hourly net load levels under the different storage scenarios.

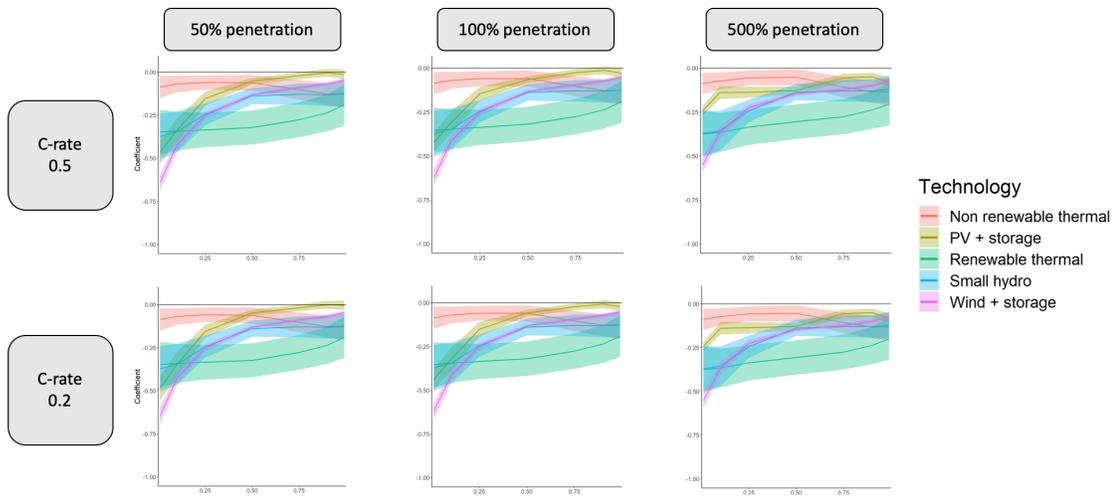


Figure 9 shows the obtained quantile impact functions for the distribution of hourly net load levels under different storage scenarios. The corresponding estimates for the highest quantiles are reported in Table 5. These results are obtained by using the same empirical strategy as before while using the quantiles of the

<sup>34</sup>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 load levels.

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

distribution of substation net load levels *after storage operations* as dependent variables. We find that very substantial levels of storage penetration would be needed to improve significantly the grid deferral potential of distributed wind and solar generation.

At 50% penetration, adding storage to distributed wind and solar investments barely increases the magnitude of their 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 are able to decrease significantly substation peak hourly net load levels, and would thus prove helpful to defer future grid expansions. Distributed PV barely starts being able to reduce substation peak net load at 100% penetration of battery storage. At 500% penetration, battery storage also shows some complementarity with distributed wind. The impact of a 1 MW investment in distributed wind bundled with storage on the 99th percentile of the distribution of hourly net load levels is twice as big as it was in the absence of

storage. The quantile impact function for wind however does not flatten as much with storage as the quantile impact function for distributed solar does. The intuition behind this observation is that, by contrast to the daily pattern of solar output, wind generation can fluctuate over longer time scales, with several days of very high or very low production levels. Storage systems with low C-rates, that is with the ability to store large amounts of energy, are thus needed to fully absorb these variations. However, because we defined storage penetration in terms of energy (see above), even higher penetration levels of storage with a low C-rate would be needed to flatten further the quantile impact function of distributed wind.

### **7.3 External validity**

Because our data covers the vast majority of both distribution substations and distributed generation capacity in France, our results provide a comprehensive assessment of the impact that investments in different distributed generation technologies had on both the net load and the ramp duration curves at distribution substations throughout France. Whether our results carry over to other countries is an open question. One potential issue is that France has its annual peak load in the winter due to its high reliance on electric heating. By contrast, a number of countries and U.S. regions experience their peak electricity load in the summer, due to a large demand for cooling.

There are however reasons to believe that our results would carry over to a signif-

icant extent to these electricity systems. In particular, although distributed solar will of course have a non-zero impact on the highest quantiles of the annual distribution of hourly net load levels 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 found that distributed wind has a small impact on the highest quantiles of the distribution of hourly net load levels even though wind output is highest during the winter. In addition, the impact of distributed wind investments on peak hourly net load levels 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, 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. Further research applying our empirical strategy to other power systems would however be needed to confirm the external validity of our results.

## **8 Conclusion**

Public policies supporting renewable electricity generation typically favor distributed units over grid-scale installations despite their significantly higher levelized cost of

generating electricity. This paper presents empirical evidence that, at least for the case of France, the 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 systems. We study the case of France, where over 28 GW of distributed generation capacity was connected to the electricity grid as of 2018. Our analysis combines two unique datasets. First, we observe the hourly net load levels 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 1 MW increase in the installed capacity of the different distributed generation technologies on both the load duration curve and the ramp duration curve supplied by substations.

Distributed wind and PV investments are found (i) to have little to no impact on the highest quantiles of the distribution of hourly net load levels, (ii) to induce large downward shifts in the lowest quantiles of the distribution of hourly net load levels, 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 significantly larger than current rates are required.

## 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 load 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

- Borenstein, Severin**, “The market value and cost of solar photovoltaic electricity production,” 2008.
- 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.
- Burger, Scott P, Jesse D Jenkins, Samuel C Huntington, and Ignacio J Perez-Arriaga**, “Why Distributed?: A critical review of the tradeoffs between centralized and decentralized resources,” *IEEE Power and Energy Magazine*, 2019, 17 (2), 16–24.
- 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.
- de Lagarde, Cyril**, “Network connection schemes for renewable energy in France: a spatial analysis,” *CEEM Working paper*, 2018.
- 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.
- Green, Richard and Iain Staffell**, ““Prosumage” and the British electricity market,” *Economics of Energy & Environmental Policy*, 2017, 6 (1), 33–50.
- 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 2019,” Technical Report 2020.
- 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.
- 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/>.

- 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.
- RTE**, “Electricity report 2018,” Technical Report 2018.
- Sexton, Steven, A Justin Kirkpatrick, Robert I Harris, and Nicholas Z Muller**, “Heterogeneous Solar Capacity Benefits, Appropriability, and the Costs of Sub-optimal 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.

## **For Online Publication - Appendices**

### **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 indeed divided into over 30,000 administrative counties.<sup>35</sup> 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.<sup>36</sup> 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

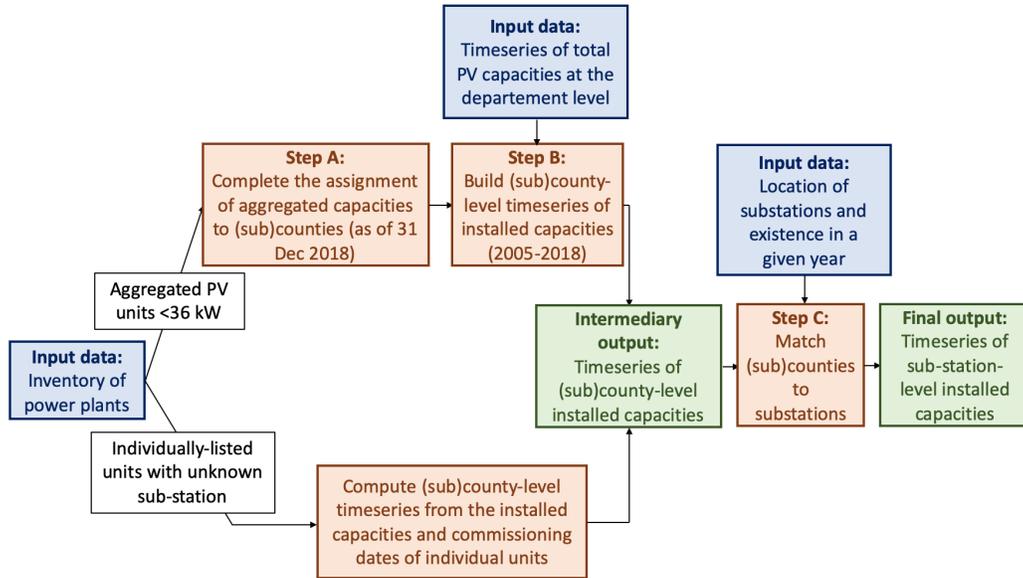
---

<sup>35</sup>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.

<sup>36</sup>The corresponding spatial boundaries were downloaded from: <https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html#contoursiris>

of mainland France into 45,508 spatial units, with a mean surface of 11.9 km<sup>2</sup> (4.6 miles squared).

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



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 timeseries of installed capacities at the (sub)county level for these units.<sup>37</sup> By contrast, observations consisting in aggregated PV units raise two challenges,<sup>38</sup> 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

<sup>37</sup>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.

<sup>38</sup>Due to their very small installed capacities, we neglect aggregated observations for technologies other than PV.

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**

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: departement, county and sub-county.<sup>39</sup> 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 departement level. Mainland France has 94 such departements.

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

---

<sup>39</sup>Because of idiosyncrasies such as mistakes when entering the fuel type of an installation, a handful of observations are aggregated at the regional level, which is a coarser spatial unit than departements. These observations however add up to less than 1 MW and are thus neglected.

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 departement 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 departement level uniformly across the pool of candidate counties where the corresponding units may be located.<sup>40</sup> 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 departement level to (sub)counties would require additional modeling and/or information sources. They are however very unlikely to affect dramatically the obtained outcome. For example, allocating capacities aggregated at the departement level to the remaining (sub)counties using log-population instead of uniform weight yields very similar results. The correlation between the

---

<sup>40</sup>For the 5% of counties that are further divided in sub-counties, two situations may arise. First, all sub-counties may each have more than 10 installations, or a total of more than 10 installations may exist in sub-counties that have less than 10 installations each. Second, less than 10 installations in total may exist in sub-counties that have less than 10 installations each. Some installations aggregated at the departement level may then be located in the latter counties, but not in the former counties.

(sub)county-level capacities obtained using uniform vs log-population weights is 0.999 (0.97 when focusing on the subset of (sub)counties where no aggregated observation is directly observed in the inventory of power plants). As a result, we use a uniform allocation for the sake of simplicity.

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

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

First, we use a third dataset from the French Department of Energy (DOE) that provides panel data at the departement level of total installed PV capacities between 2006 and 2018 (installed capacities being virtually zero in 2005).<sup>41</sup> 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

---

<sup>41</sup>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](http://www.statistiques.developpement-durable.gouv.fr/tableau-de-bord-solaire-photovoltaique-quatrieme-trimestre-2018) for the fourth quarter of 2018). We are grateful to the Department of Energy for having shared with us the corresponding historical data (updated of as of July 2020).

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 timeseries of PV capacity from individually-listed units. Subtracting these timeseries to the timeseries of department-level total PV capacity from the DOE dataset yields department-level timeseries of PV capacity from aggregated units. Figures A.3 and A.4 show the obtained 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 timeseries. We further impose monotonicity which is (mildly) violated on only three occasions.

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).

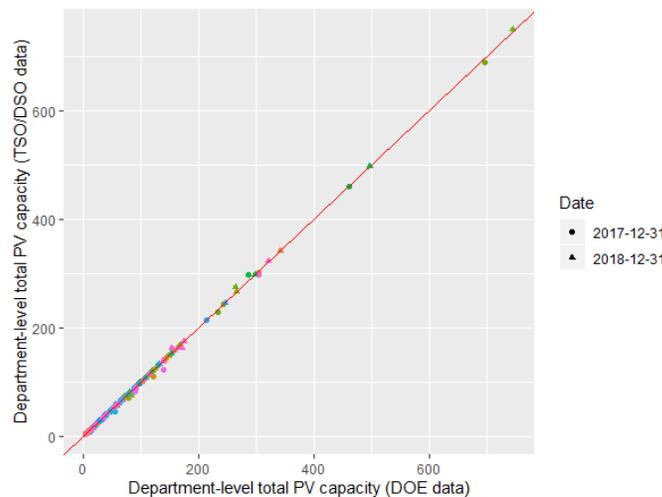
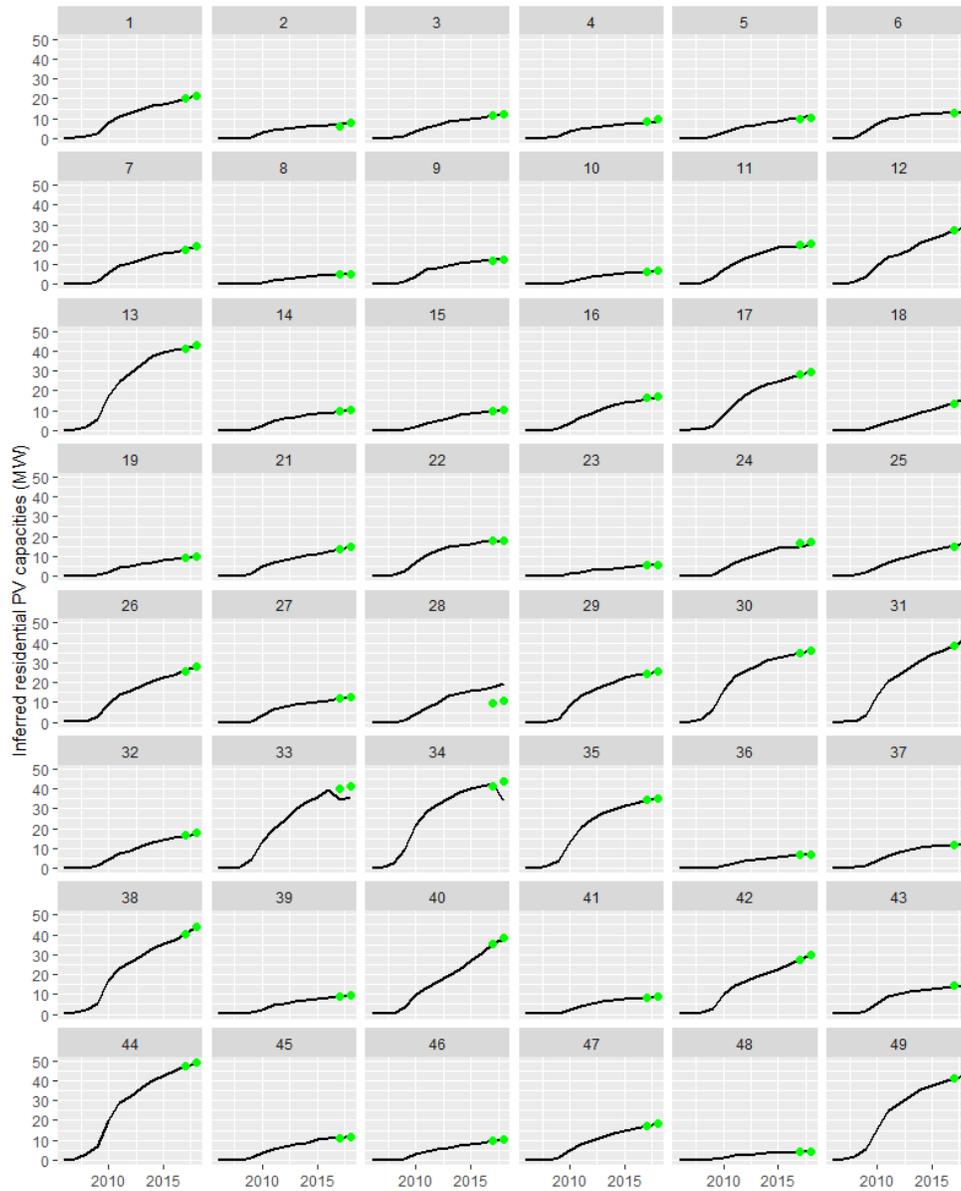
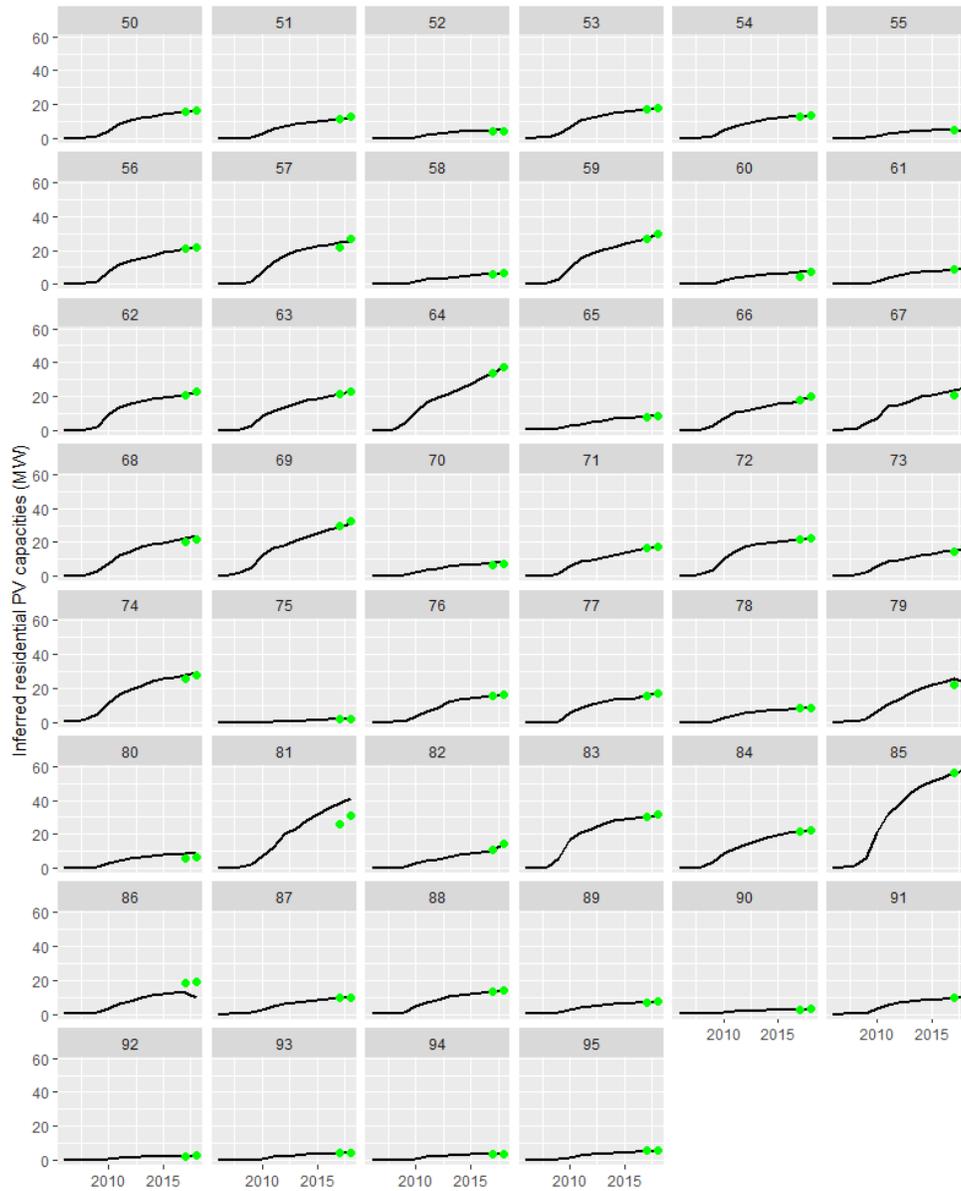


Figure A.3: Inferred department-level timeseries of PV capacity from aggregated units (first 48 departements). Green dots represent actual capacities as reported in the public inventories of power plants of 2017 and 2018.



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.

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



In other words, we want to define capacities  $K_{c,d,y}$  for each year  $y$  and (sub)county

$c$  (located in departement  $d$ ) such that:

$$\forall y, \forall d, \sum_{c \in d} K_{c,d,y} = K_{d,y} \quad (3)$$

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

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

$$K_{c,d,y}^{HS} \equiv \frac{W_{c,d}}{\sum_{c' \in d} W_{c',d}} K_{d,y} \quad (4)$$

2. **Sequential static approach:** this method considers that new PV units get installed first in the counties with the highest remaining capacity to be installed. In other words, knowing that  $W_{c,d}$  must be installed by 2018,<sup>42</sup> we compute  $K_{d,y}^*$  such that:

$$\sum_{c \in d} \max(W_{c,d} - K_{d,y}^*, 0) \equiv K_{d,y} \quad (5)$$

We then postulate that the installed capacity in county  $c$  of departement  $d$  as

---

<sup>42</sup>More precisely,  $W_{c,d}$  is normalized within each departement in order to sum to  $K_{d,2018}$  and thus be more consistent with the rest of the timeseries. As shown on Figures A.3 and A.4,  $K_{d,2018}$  and  $\sum_{c \in d} W_{c,d}$  are virtually identical for almost every departement.

of year  $y$  was:

$$K_{c,d,y}^{SS} \equiv \max(W_{c,d} - K_{d,y}^*, 0) \quad (6)$$

3. **Homothetic dynamic approach:** although their exact meaning is somewhat ambiguous, the public inventory of power plants does provide “commissioning dates” for aggregated PV observations. We interpret 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  $y$ .<sup>43</sup> We then define the installed capacity in county  $c$  of department  $d$  as of year  $y$  using an iterative approach:

$$K_{c,d,y}^{HD} \equiv \mathbf{1}_{c \in \mathcal{C}_{d,y}} \left( K_{c,d,y-1}^{HD} + \frac{W_{c,d} - K_{c,d,y-1}^{HD}}{\sum_{c' \in \mathcal{C}_{d,y}} W_{c',d} - K_{c',d,y-1}^{HD}} (K_{d,y} - K_{d,y-1}) \right) \quad (7)$$

where  $K_{c,d,2005}^{HD} = 0$  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

---

<sup>43</sup>(Sub)counties for which no aggregated observation exist in the inventory are assumed to belong to  $\mathcal{C}_{d,y}$  for all  $y$ .

compute  $K_{d,y}^*$  such that:

$$\sum_{c \in \mathcal{C}_{d,y}} \max(W_{c,d} - K_{c,d,y-1}^{SD} - K_{d,y}^*, 0) \equiv K_{d,y} - K_{d,y-1} \quad (8)$$

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

$$K_{c,d,y}^{SD} \equiv \mathbf{1}_{c \in \mathcal{C}_{d,y}} (K_{c,d,y-1}^{SD} + \max(W_{c,d} - K_{c,d,y-1}^{SD} - K_{d,y}^*, 0)) \quad (9)$$

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

In order to match (sub)counties to substations, we rely on two sources of information. First, we know the GPS coordinates of the substations.<sup>44</sup> Second, we observe

---

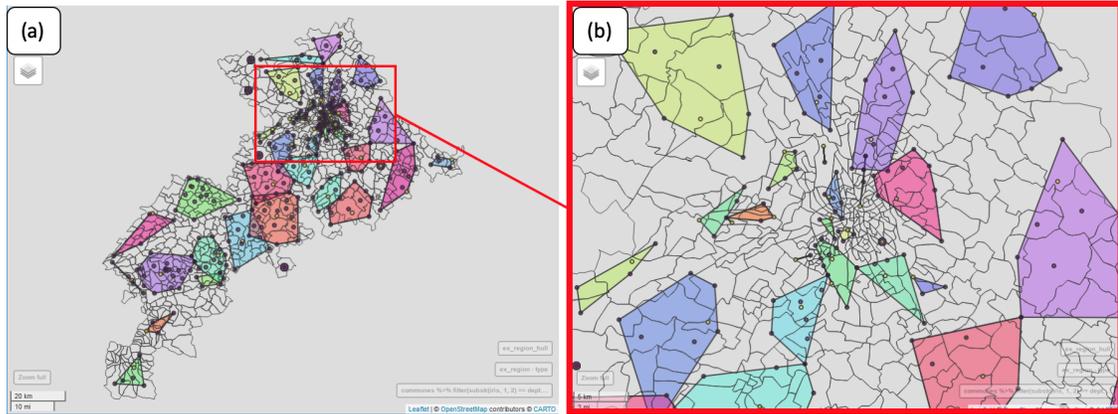
<sup>44</sup>This information is for example available from: [www.data.gouv.fr/en/datasets/postes-electriques-rte-au-6-juin-2020-1/](http://www.data.gouv.fr/en/datasets/postes-electriques-rte-au-6-juin-2020-1/) (last accessed on 31 August 2020).

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 départements. Panel (b) further zooms on a densely populated area where we further divided counties into sub-counties. Even in urban areas, the spatial units we use appears to be granular enough relative to the spatial density of distribution substations.

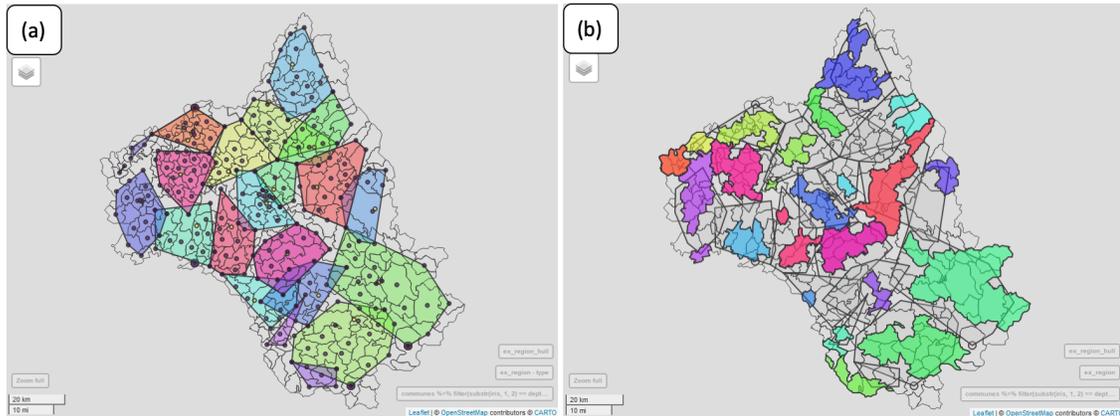
We then use the computed convex hulls and the knowledge of the spatial boundaries of (sub)counties to build a mapping from (sub)counties to substations. First, a (sub)county  $c$  that intersects with the convex hull of substation  $s$  is assumed to connect to this substation. If a given (sub)county intersects with several convex hulls, a distributed generation unit located in this (sub)county is deemed equally likely to connect to the corresponding substations. This first step maps almost two thirds of

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



our spatial units (29,330 out of 45,508). In addition, over half of our spatial units (24,585 out of 45,508) intersect with a single convex hull. Figure A.6 illustrate this first step by showing, for one of the 94 departements, the counties that intersect with a single substation convex hull. In a second step, we isolate remaining (sub)counties that are adjacent to one or several (sub)counties that were all matched in step 1 to the same substation. These (sub)counties are assumed to also connect to the corresponding substation. This second step, which maps 6,523 additional spatial units 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): obtained convex hulls for the Aveyron departement. 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.

## **B Sensitivity Analyses**

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

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

As discussed in Appendix A, 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 departements. 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 departement-level capacities.

The timeseries 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.

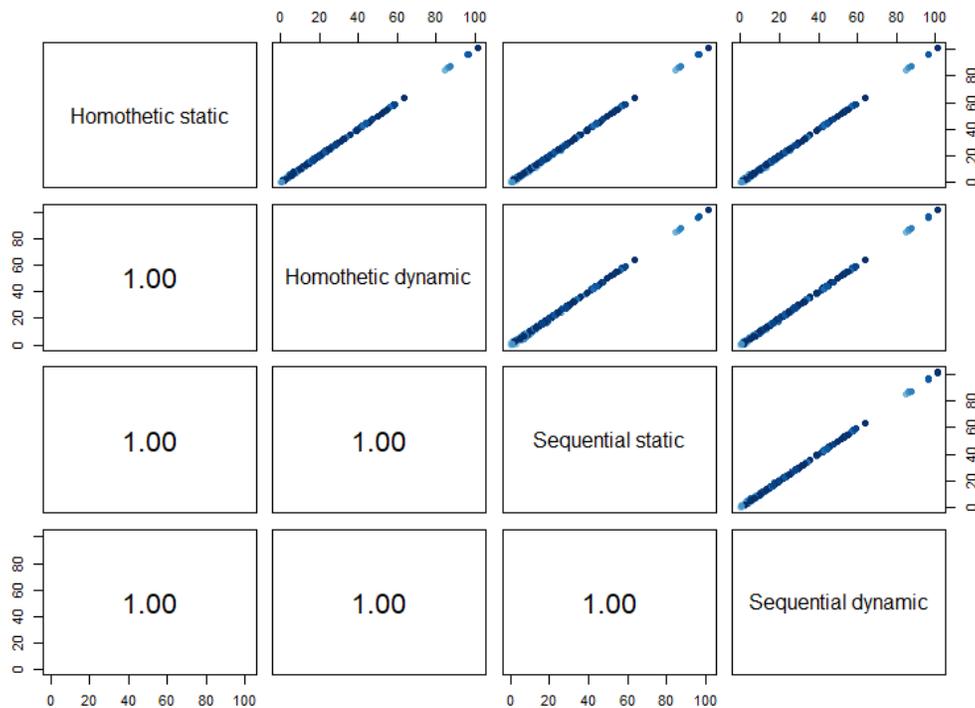


Figure B.1 shows that the four PV metrics we obtain are virtually identical. This

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:

$$K_{t,s,y} \equiv K_{t,s,y}^0 + \hat{K}_{t,s,y} \quad (10)$$

where:

- $K_{t,s,y}^0$  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<sup>2</sup>) of the installed capacities of each technology between known and inferred capacities.

Technology	$\text{Var}(K_{t,s,y}^0)$	$\text{Var}(\hat{K}_{t,s,y})$	$\text{Var}(K_{t,s,y})$
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  $K_{t,s,y}^0$ ,  $\hat{K}_{t,s,y}$  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  $K_{t,s,y}^0$  instead of  $K_{t,s,y}$  as independent variables. These results are almost identical to the results reported in the main text.

## B.2 Sensitivity to our model specification

### Two-way fixed effect specification

Our econometric model is a two-way fixed effect specification. This very parsimonious approach accounts for both substation-specific unobserved characteristics

Table B.2: Estimated coefficients when regressing the main quantiles of the annual distribution of hourly net load levels on the installed capacities  $K_{t,s,y}^0$  from units whose upstream substation is known. Robust standard errors clustered at the substation level are reported.

	<i>Dependent variable:</i>						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Wind	-0.723 (0.018)	-0.463 (0.014)	-0.271 (0.008)	-0.139 (0.004)	-0.091 (0.005)	-0.066 (0.005)	-0.035 (0.006)
PV	-0.557 (0.035)	-0.384 (0.034)	-0.173 (0.017)	-0.052 (0.010)	-0.020 (0.010)	-0.0004 (0.012)	0.004 (0.014)
Small hydro	-0.394 (0.066)	-0.359 (0.061)	-0.249 (0.033)	-0.139 (0.024)	-0.126 (0.031)	-0.127 (0.034)	-0.127 (0.037)
Renewable thermal	-0.390 (0.065)	-0.373 (0.057)	-0.352 (0.051)	-0.332 (0.051)	-0.279 (0.052)	-0.234 (0.056)	-0.182 (0.062)
Non renewable thermal	-0.083 (0.034)	-0.069 (0.024)	-0.060 (0.021)	-0.065 (0.024)	-0.107 (0.029)	-0.130 (0.032)	-0.127 (0.040)
Observations	30,091	30,091	30,091	30,091	30,091	30,091	30,091
R <sup>2</sup>	0.958	0.960	0.976	0.983	0.983	0.985	0.984
Adjusted R <sup>2</sup>	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  $K_{t,s,y}^0$  from units whose upstream substation is known. Robust standard errors clustered at the substation level are reported.

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

(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 con-founders, the existence of unobserved variables that explain electricity consumption could still bias our estimates if these variables have experienced both significant and spatially contrasted changes over our period of interest, and are correlated with installed distributed generation capacities. For example, although residential PV represents a small fraction of total PV capacities,<sup>45</sup> population growth could both increase electricity consumption and be correlated with PV adoption if solar panels are more likely to be installed on the roof of new buildings. To test for this possibility, we compare the obtained results from our main specification to the results obtained when adding year-by-region (mainland France has 12 regions) and year-by-departement (mainland France has 94 departements) fixed effects. Table B.4 reports the obtained results when using mean hourly net load level as our dependent variable, which enables a direct interpretation of the estimated coefficients as capacity factors (and thus to assess their credibility). Our estimates appear to be robust to the inclusion of these additional controls. Similar conclusions were reached with specifications using the values of the different quantiles of the distribution of hourly net load levels or hourly ramps as dependent variables.

### **Interaction terms**

Our main specification does not include interaction terms between the installed

---

<sup>45</sup>Residential PV installations are included in the small aggregated PV units, and thus represent at most a quarter of total capacities.

Table B.4: Estimated coefficients when regressing mean hourly net load level on installed capacities (i) using our main specification (column (1)) ; (ii) adding year-by-region fixed effects (column (2)) ; and (iii) adding year-by-departement fixed effects (column (3)).

	<i>Dependent variable:</i>		
	Mean hourly net load level		
	(1)	(2)	(3)
Wind	-0.192 (0.006)	-0.187 (0.007)	-0.180 (0.007)
PV	-0.109 (0.012)	-0.118 (0.012)	-0.120 (0.011)
Small hydro	-0.192 (0.025)	-0.209 (0.025)	-0.200 (0.024)
Renewable thermal	-0.302 (0.050)	-0.298 (0.052)	-0.296 (0.051)
Non renewable thermal	-0.081 (0.024)	-0.082 (0.024)	-0.088 (0.026)
year FE	Y	N	N
substation FE	Y	Y	Y
year by region FE	N	Y	N
year by departement FE	N	N	Y
Observations	30,091	30,091	30,091
R <sup>2</sup>	0.983	0.984	0.985
Adjusted R <sup>2</sup>	0.982	0.983	0.983

capacities of distinct distributed generation technologies. As shown in Table B.5, 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 (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.

Table B.5: Obtained results for the load duration curve when including interactions terms. Robust standard errors clustered at the substation level are reported.

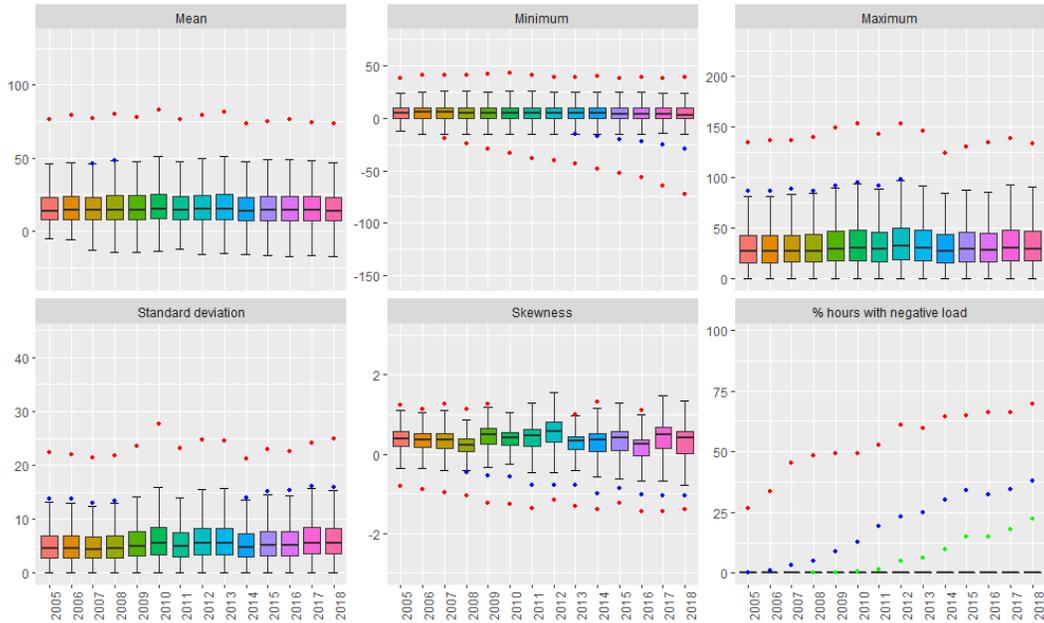
	<i>Dependent variable:</i>						
	Q1	Q10	Q25	Q50	Q75	Q90	Q99
Non_renewable_thermal	-0.100 (0.041)	-0.087 (0.032)	-0.064 (0.025)	-0.059 (0.028)	-0.104 (0.035)	-0.126 (0.038)	-0.130 (0.046)
PV	-0.546 (0.041)	-0.375 (0.040)	-0.160 (0.021)	-0.040 (0.011)	-0.013 (0.012)	0.004 (0.013)	0.011 (0.016)
Renewable_thermal	-0.389 (0.086)	-0.377 (0.073)	-0.350 (0.065)	-0.296 (0.064)	-0.245 (0.068)	-0.198 (0.078)	-0.176 (0.091)
Small_hydro	-0.398 (0.070)	-0.360 (0.064)	-0.226 (0.037)	-0.114 (0.023)	-0.113 (0.029)	-0.124 (0.033)	-0.125 (0.037)
Wind	-0.695 (0.028)	-0.440 (0.021)	-0.249 (0.012)	-0.126 (0.006)	-0.083 (0.006)	-0.061 (0.006)	-0.032 (0.007)
PV_x_Wind	0.010 (0.004)	0.005 (0.003)	0.0001 (0.002)	-0.001 (0.001)	-0.001 (0.001)	-0.0004 (0.001)	-0.001 (0.001)
PV_x_Hydro	0.018 (0.009)	0.014 (0.007)	0.0002 (0.004)	-0.005 (0.003)	-0.0004 (0.002)	0.002 (0.003)	0.001 (0.004)
PV_x_Ren_thermal	-0.0005 (0.011)	0.001 (0.009)	0.0004 (0.006)	-0.004 (0.005)	-0.005 (0.004)	-0.005 (0.004)	-0.004 (0.006)
PV_x_Non_Ren_thermal	0.006 (0.010)	0.008 (0.008)	0.004 (0.004)	0.002 (0.003)	0.003 (0.004)	0.002 (0.004)	0.002 (0.005)
Wind_x_Hydro	-0.023 (0.017)	-0.021 (0.013)	-0.018 (0.009)	-0.013 (0.006)	-0.009 (0.004)	-0.007 (0.004)	-0.007 (0.006)
Wind_x_Ren_thermal	0.007 (0.011)	0.005 (0.008)	0.004 (0.005)	0.0001 (0.004)	-0.004 (0.004)	-0.004 (0.004)	-0.004 (0.004)
Wind_x_Non_Ren_thermal	-0.0004 (0.004)	0.0005 (0.004)	0.0004 (0.002)	0.001 (0.001)	0.001 (0.002)	0.001 (0.002)	0.002 (0.002)
Hydro_x_Ren_thermal	0.040 (0.041)	0.049 (0.037)	0.039 (0.033)	0.025 (0.035)	0.030 (0.041)	0.030 (0.047)	0.046 (0.057)
Hydro_x_Non_Ren_thermal	-0.003 (0.012)	-0.006 (0.012)	-0.010 (0.012)	-0.012 (0.012)	-0.016 (0.013)	-0.013 (0.014)	-0.005 (0.016)
Ren_thermal_x_Non_Ren_thermal	0.001 (0.008)	-0.001 (0.006)	-0.003 (0.006)	-0.004 (0.005)	-0.002 (0.006)	-0.004 (0.006)	0.0003 (0.007)
Observations	30,091	30,091	30,091	30,091	30,091	30,091	30,091
R <sup>2</sup>	0.953	0.958	0.975	0.983	0.984	0.985	0.984
Adjusted R <sup>2</sup>	0.950	0.955	0.973	0.981	0.982	0.983	0.983

## 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 load level for substation  $s$  during hour  $h$  of year  $y$ . Various summary statistics can then be derived from these distributions. Figure C.2 shows the evolution, between 2005 and 2018, of the cross-sectional distribution of the following summary statistics: mean, minimum, maximum, standard deviation, skewness, and percentage of hours with negative net load, that is hours during which power was flowing from the distribution to the transmission grid.

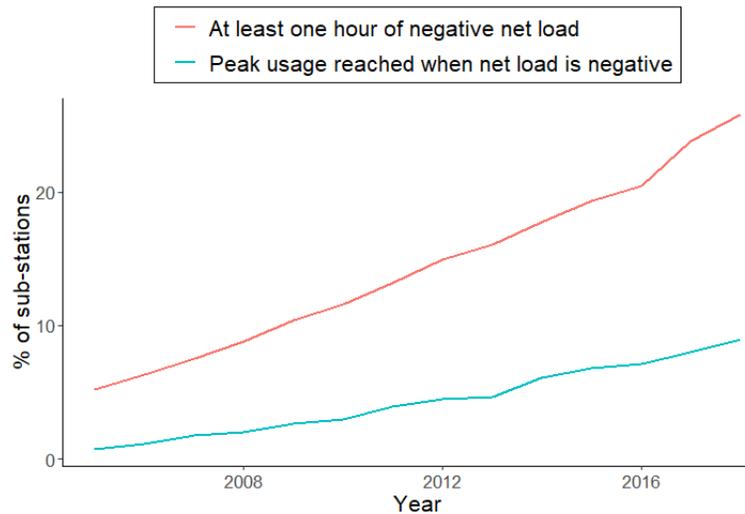
We observe that the summary statistics that exhibit the most significant changes relate to reverse power flows, that is to hours during which local generation exceeds local consumption. For example, the left tail of the distribution of minimum net load and the right tail of the distribution of the percentage of hours with negative net load have expanded significantly between 2005 and 2018. Figure C.3 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 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 can also estimate the impact that different distributed generation technologies have had on the summary statistics plotted on Figure C.2. To do so, we use the specification of Equation (1) using these summary statistics as dependent variables. Table C.6 shows the obtained 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

Figure C.3: 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.



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 (2018)). Power losses between the generation site and the substation, as well as the fact that installed capacities are measured as of 31 December, are possible 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.<sup>46</sup> Second, we observe that different technologies have very contrasted impacts on minimum and maximum net load. While non-renewable thermal units impact minimum and

<sup>46</sup>Consistently, estimating the same model when replacing total installed PV capacity by installed PV capacity from units for which we observe the upstream substation (which tend to be larger installations) yields a capacity factor of 12%.

maximum net load in a similar way as 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. By contrast, thermal units have much milder impacts. Finally, we observe that reverse power flows seem to be driven by wind, PV and small hydro capacities.

Table C.6: 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.

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

## 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 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  $\tau$  is then the optimized value of the following problem:

$$\begin{aligned} \tau &\equiv \min_{\delta} (\hat{\beta} - \delta)^T \text{Var}(\hat{\beta})^{-1} (\hat{\beta} - \delta) \\ &\text{s.t.} \\ &\text{HC0-h} \end{aligned} \tag{11}$$

The constraint HC0-h formalizes the different null hypotheses we test in terms of linear equality or inequality constraints on  $\delta$ . 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.7 reports the obtained test statistics.

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

Table D.7: Obtained test statistics

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

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  $\Pr[\chi_1^2 \geq 2.706] = 0.1$ , we cannot reject the null hypothesis even at the 0.1 level whenever the test statistic is lower than 2.706. Conversely, since  $\Pr[\chi_6^2 \geq 16.812] = 0.01$ , 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  $\chi^2$  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.

## E Quadratic Specification Results

This Appendix reports the obtained results with a quadratic specification in installed capacities of distributed generation. Figures E.4 and E.5 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.8: 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.

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

Table E.9: 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.

	<i>Dependent variable:</i>						
	Q01	Q10	Q25	Q50	Q75	Q90	Q99
PV	-0.071 (0.017)	-0.026 (0.004)	-0.017 (0.001)	-0.003 (0.001)	0.017 (0.001)	0.037 (0.005)	0.054 (0.015)
PV <sup>2</sup>	-0.002 (0.001)	-0.001 (0.0001)	0.00003 (0.00002)	0.00001 (0.00001)	0.00005 (0.00002)	0.001 (0.0002)	0.002 (0.001)
Wind	-0.136 (0.012)	-0.040 (0.004)	-0.016 (0.002)	0.0001 (0.0001)	0.017 (0.002)	0.041 (0.004)	0.130 (0.012)
Wind <sup>2</sup>	-0.0001 (0.0002)	-0.0001 (0.0001)	-0.00001 (0.00002)	0.00000 (0.00000)	0.00001 (0.00002)	0.0001 (0.0001)	0.0002 (0.0002)
Small_hydro	0.004 (0.022)	-0.0004 (0.006)	0.005 (0.003)	0.0004 (0.002)	-0.002 (0.003)	-0.013 (0.007)	0.011 (0.030)
Small_hydro <sup>2</sup>	-0.001 (0.001)	-0.0001 (0.0002)	-0.0002 (0.0001)	0.00002 (0.0001)	0.0001 (0.0001)	0.0005 (0.0002)	0.001 (0.001)
Renewable_thermal	-0.017 (0.020)	-0.006 (0.009)	-0.001 (0.004)	-0.001 (0.002)	0.002 (0.005)	0.003 (0.010)	0.032 (0.021)
Renewable_thermal <sup>2</sup>	0.00003 (0.001)	0.0003 (0.001)	0.0001 (0.0003)	0.0002 (0.0001)	-0.0001 (0.0004)	-0.0003 (0.001)	-0.002 (0.002)
Non_renewable_thermal	-0.008 (0.012)	0.002 (0.005)	0.002 (0.002)	-0.0004 (0.001)	-0.004 (0.002)	-0.001 (0.005)	0.007 (0.013)
Non_renewable_thermal <sup>2</sup>	0.0003 (0.0004)	-0.00004 (0.0002)	-0.0001 (0.0001)	-0.00004 (0.00004)	0.0002 (0.0001)	0.0001 (0.0002)	-0.0003 (0.0004)

Figure E.4: 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)

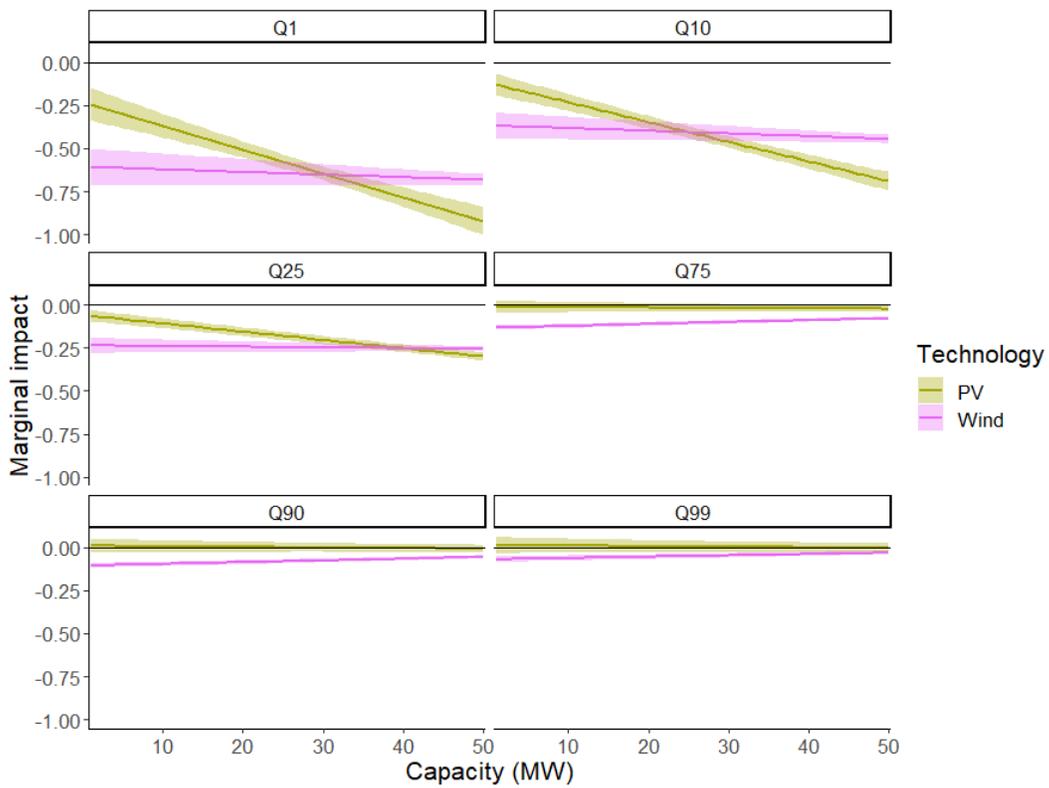
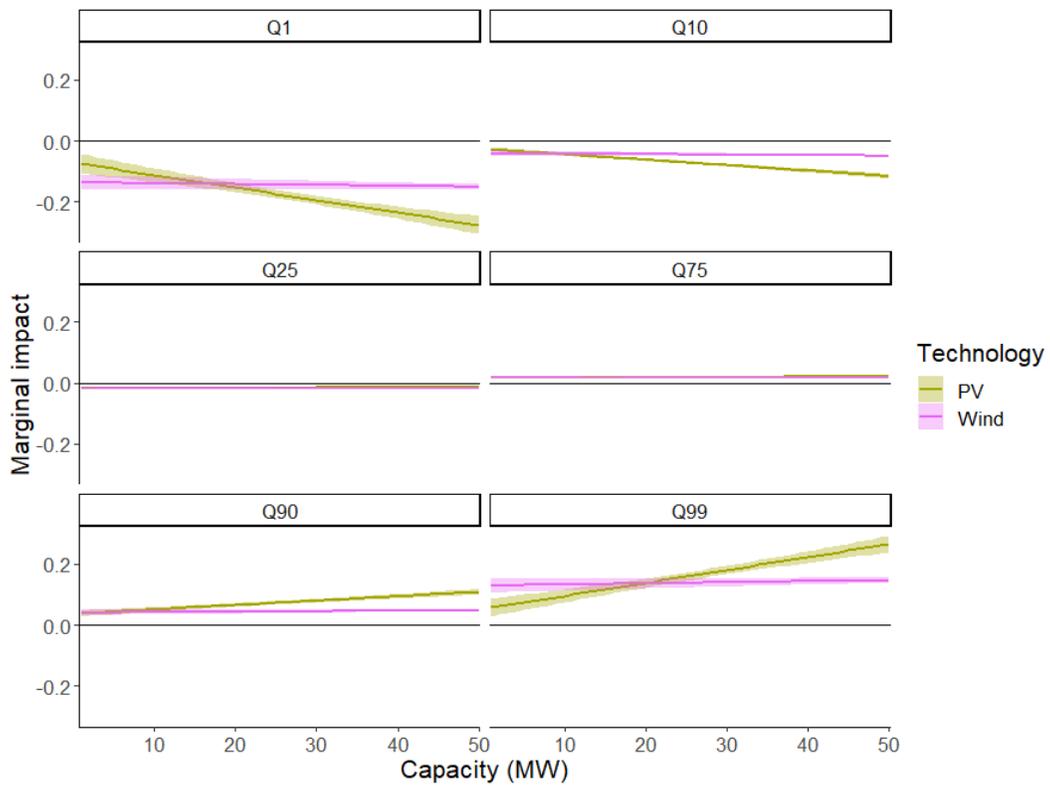


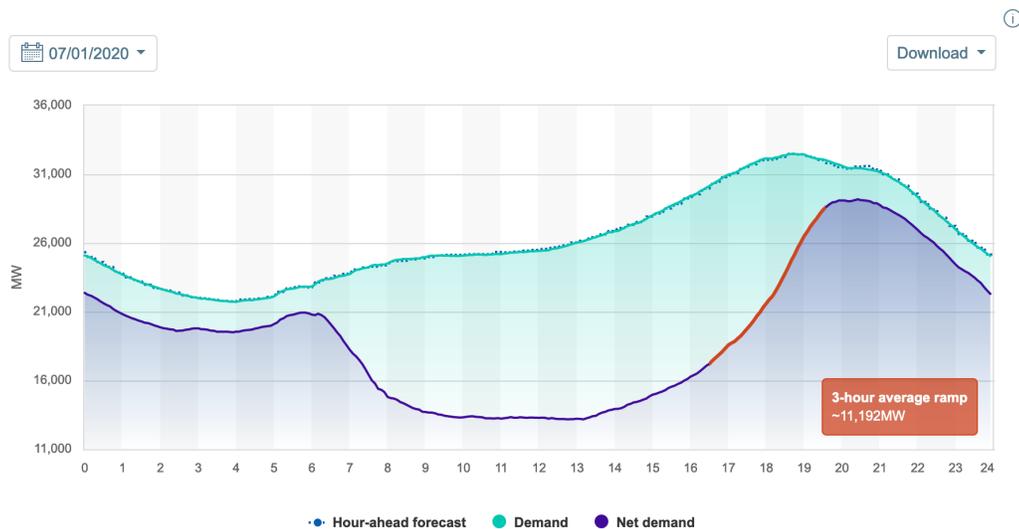
Figure E.5: Marginal impact of wind and solar on the ramp duration curve as a function of installed capacity (confidence intervals are built from the variance-covariance matrix with errors clustered at the sub-station level)



## F Discussion of External Validity

Figure F.6 shows demand, along with demand net of wind and solar generation, over CAISO territory on 1 July 2020. Although wind and solar produce very significant amounts of energy over the course of the day, PV production decreases progressively in the evening and stops completely at night. The resulting peak net demand is shifted later in the evening when PV output is negligible, so that the difference between peak gross demand and peak net demand only amounts to 3.3 GW despite an average output from wind and solar for the day of 7 GW.

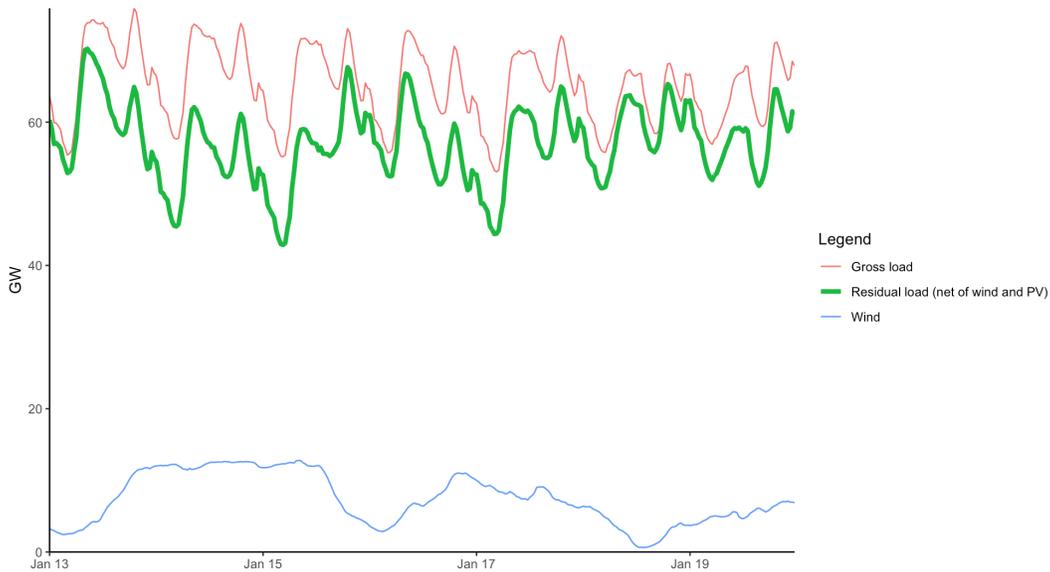
Figure F.6: CAISO hourly gross and net demand for 1 July 2020 (screenshot taken from <http://www.caiso.com/TodaysOutlook/Pages/default.aspx>)



Similarly, Figure F.7 plots demand, along with demand net of wind and solar generation, in France for the week from 13 January to 19 January 2020. Similarly, we observe that peak net demand is shifted to hours with low wind output. As a re-

sult, the difference between peak gross demand and peak net demand only amounts to 5.7 GW although the average output from wind and solar for the week is 8 GW.

Figure F.7: National gross and net hourly demand in France for the week of 13 January to 19 January 2020, along with total hourly wind output (from both distributed and utility-scale units). Corresponding data was downloaded from RTE (<https://www.rte-france.com/eco2mix>)



Overall, these observations suggest that, for both France and California, the high level of contemporaneous correlation in the output of wind/solar units mechanically creates hours with consistently low wind/solar generation. 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. As a result, it seems likely that many of our results for France would carry over at least partially to summer-peak systems like California.