Quantifying Resolution Implications for Agent-based Distributed Energy Resource Customer Adoption Models

Ted Kwasnik, Benjamin Sigrin, and David Bielen

National Renewable Energy Laboratory

Quantifying Resolution Implications for Agent-based Distributed Energy Resource Customer Adoption Models

Ted Kwasnik, Benjamin Sigrin, and David Bielen

National Renewable Energy Laboratory

NOTICE

Acknowledgments

The authors thank Maura Bardos, Erin Boedecker, David Kinyon, John Maples, Brian Bush and Mike Meshek for their thoughtful reviews.

List of Acronyms

ABM	Agent-based model
CBECS	Commercial Buildings Energy Consumption Survey
DPV	Distributed photovoltaics
NREL	National Renewable Energy Laboratory
RECS	Residential Energy Consumption Survey

Executive Summary

Distributed photovoltaics (DPV) are a growing source of electricity generation in the United States, and with adoption driven by customer behavior and localized economics, projecting the deployment of this technology is a challenging analytical problem. Moreover, understanding the sources of uncertainty in customer adoption models and how they can be reduced is important to a range of stakeholders that use their outputs, including grid planners, regulators, and industry. Most prior studies have used top-down methods, such as the use of population central tendencies to project aggregate adoption. In contrast, a growing field of work seeks to use bottom-up methods (i.e., individual-level decision-making).

We explore trade-offs of top-down and bottom-up methods in their precision and computational burden using the National Renewable Energy Laboratory's (NREL's) Distributed Generation Market Demand (dGen) model, an agent-based model of residential and nonresidential distributed PV adoption. In particular, we assess the role of agent resolution in instantiating statistically-representative populations in the model—and the resulting variance of model projections at the state, sector, and county levels. At low sampling rates, the model resembles a top-down model, whereas as sampling rates increase dGen converges to a bottom-up structure by simulating more unique customer types. Though sampling-based models such as dGen can be operated with many agents to ensure accuracy, doing so greatly increases the computational burden of the simulation. This report lends insight into whether high-resolution results can be approximated sufficiently well using fewer computational resources.

At the state and sector levels, we find systematic differences in mean projected DPV capacity adopted across different levels of agent resolution, namely that cumulative DPV adoption estimates decrease with greater agent resolution. Following this, we conclude that sets of low agent runs in dGen cannot be directly substituted for high-resolution runs. Moreover, we find that the average size of installed DPV explains decreases in system sizes by agent resolution, and we attribute this trend to a system-sizing mechanism that biases smaller sizes when agent annual load and roof area suitable to PV are not positively correlated. Thus, associating suitable roof area with annual load on an agent by agent basis in a manner more consistent with the relationship inherent in input data sets is expected to improve the interpretability of results. At the county level, we generally see similar trends to the state and sector levels; however, we also note the rates at which DPV estimates actually increase with agent resolution.

We also find that the variance in projected DPV capacity adoption decreases substantially as agent resolution increases. This trend holds at the state and sector levels and for most cases at the county level. We find that precision improvements in both sectors are greatest at low levels of agent resolution, which matches our intuition that estimates will differ less across model runs as underlying samples increasingly resemble the full population. We also find reason to limit agent resolution in that each additional agent adds about 3 minutes to a minimum solve time of 18 minutes. Also, significantly fewer than 100 iterations may be sufficient to achieve acceptable levels of precision; state-level variance varies by less than 5% beyond about 50 runs relative to the average of 100 runs. At the county level, we can explain most of the variance in the nonresidential sector, yet this scale reveals trends that diverge from those observed at the state and sector levels that are not fully explained. Therefore, we encourage further research into sources of variance in both sectors at this more-refined resolution.

Additionally, we observe that while the mean and variance trends are similar for the residential and nonresidential sectors, the magnitude of the variance for the nonresidential sector is substantially larger. This suggests modeling precision could be improved by increasing the number of agents sampled in the nonresidential sector. For example, extrapolating empirical results suggests that while 20-agent resolution is sufficient to settle on a standard deviation that is 1% of the mean in the residential sector, more than two and a half times as many agents are needed in the nonresidential sector to achieve the same level of precision. To date, there is little published agent-based modeling of nonresidential DPV adoption, and the variance observed in this model suggests more research is needed to better capture the complexities introduced by this sector's more varied building types and customer use cases.

Important further research will explore how key agent resolution trends we observed here play out in new geographies. As new agent-level attributes are added (e.g., socioeconomic status and environmental concern), we will also want to reassess these trends for consistency. Beyond agent resolution, understanding variance sensitivity to additional agent and county characteristics, such as DPV market maturity and macroeconomic parameters (e.g., technology costs and fuel prices) will improve the interpretation of results. Agent-based models have an increasingly important role to play in grid planning and policy analysis, and explorations such as this one into their mechanics will help inform critical decisions with increased certainty.

Table of Contents

List of Figures

List of Tables

1 Introduction

Customer-adopted distributed solar photovoltaics (DPV) are a growing source of electricity generation capacity in the United States. Over one gigawatt was added to national DPV capacity in the first quarter of 2018 [1], and recent trends suggest cumulative installed capacity may nearly double by 2023 [2].

Projecting future customer adoption of new distributed energy resources at high spatial resolution is recognized as a key challenge in the power systems domain. Unlike utility-scale generation assets, which are procured directly by centralized utilities or independent power producers, the adoption of DPV is driven by the purchasing decisions of broad customer bases, both residential and nonresidential (i.e., commercial and industrial). Accordingly, its adoption is subject to highly localized factors, including the local incentives and levels of neighborhood adoption [3], and its diffusion is not uniform across populations. Grid planners and researchers often turn to data-driven models to provide insight into the key drivers of future DPV diffusion and resulting changes in the electricity sector. As this question of DPV adoption poses a challenging analytical problem, effective use of the models requires a keen understanding of sources of uncertainty in their inputs and algorithms.

In this report, we investigate how different population-sampling strategies impact the distribution of technology adoption in the context of an agent-based model (ABM). Specifically, we employ NREL's dGen model, which samples one or more agents in each U.S. county before simulating their DPV adoption decisions over time to better understand the sensitivity of long-term estimates of DPV adoption to agent resolution. Our analysis examines DPV capacity at the state, sector, and county levels across iterative stochastic simulations at four agent resolutions to assess trends in the mean and variance. We also assess whether sets of low-agent simulations can be used to simulate high-resolution runs. Finally, we report observed model run times, precision variation by number of iterative simulations, and extrapolated trends in precision by agent resolution to inform model calibration decisions.

The results of this analysis provide insight into the trade-off between computational expense and model accuracy. Though sampling-based models such as dGen can be operated with many agents to ensure accuracy, doing so greatly increases the computational burden of the simulation. Burdensome computing can result in prolonged run times or even the inability to complete a simulation.

This report lends insight into whether high-resolution results can be approximated sufficiently well using fewer computational resources. It also informs the interpretation of dGen results and the design of stochastic sampling processes within similar models at a fundamental level. We anticipate that the empirical process we use to investigate the dGen model will help inform best practices regarding the calibration and validation of other stochastic ABMs.

The report also contributes to a growing literature on methods for the projection of DPV customer adoption. These methods differ widely in the algorithms and input data used, though they can generally be classified as either top-down or bottom-up. Top-down models use central tendencies to project aggregate deployment, notable examples include time series models [4, 5] and econometric models [6–10]. Bass Models are perhaps the most widely used adoption models

[5 –11] . They capture the aggregate decisions of a heterogeneous population of technology innovators and imitators, which generally cannot be localized at high spatial resolutions.

In contrast, bottom-up adoption models seek to simulate the decision-making of individual consumers, or at least the decisions of clusters of statistically-similar ones. They are an alternative to top-down methods in that they can simulate the decision-making processes of individual agents within heterogeneous populations and can account for external constraints and decision drivers that vary geographically at a high spatial resolution. And, bottom-up adoption models capture the effect of social influence.

Agent-based modeling (ABM) is increasingly employed to project DPV adoption [5, 16–25] and is emerging as a critical method to inform distribution grid planning. Alyousef et al. [17] found that income, payback period, maximum budget, degree of environmental concern, and social influence largely explained the decision to adopt DPV or not in their German study population. Similarly, Rai and Robinson [16] calibrated agents using geographic data sets and customer surveys in Austin, Texas, using the theory of planned behavior [26] to simulate DPV adoption choices. NREL's dGen model [27], and its predecessor SolarDS [19], take a hybrid Bass Model and ABM approach wherein synthetic spatially-aware agents are instantiated with characteristics reflective of their surroundings, and their adoption patterns are fitted to a Bass Model logistic curve based on prevailing economic trends.

Several inherent sources of uncertainty are associated with agent-based DPV models. For example, high-resolution empirical data are commonly not available for study populations, so representative profiles or stochastic sampling from distributions are employed to simulate agents' characteristics (e.g., income, building type and size, hourly load consumption) and geographic contexts (e.g., solar insolation and utility rate structure). Moreover, exogenous market parameters (e.g., inflation and capital costs) are also commonly used, based on assumptions about sustained patterns over long periods. However, few studies report on the sensitivity of results to agent resolution or how estimates vary by the number of stochastic model runs, which is the principal topic of this analysis.[1]

The rest of the report is organized as follows. Section 2 describes the dGen model and lays out the study methodology. The results are presented in Section 3. They are followed by a discussion in Section 4 and conclusions Section 5.

---

[1] Drury et al. (2014) quantify DPV modeling uncertainty in regards to investment risk, and they find that while inter-annual solar variability at a given location tends to cancel itself out over multiple years, PV technical performance and variable retail electricity markets pose significant sources of modeling uncertainty [28].

2 Methodology

This analysis reported here employs NREL's dGen model,[2] an agent-based simulation platform that projects customer DPV adoption trends from 2014 to 2050 [29]. At the center of this analysis is a conceptual abstraction of the dGen model as a multivariate stochastic function for which the distribution of the output is unknown and will be estimated from a set of model runs. Because dGen assimilates disparate data sets based on interdependent relationships and conditional logic, the structure of underlying sampled distributions for attributes varies from agent to agent and can be burdensome to isolate programmatically. Each agent in effect may be endowed with attributes sampled from unique distributions as a result of pre-established traits (i.e., building size class and building type). Based on the law of large numbers, however, we propose that the outputs of dGen at the state level will display a central tendency over enough iterations, given fixed parametrization.

For this analysis, we run dGen repeatedly at varying agent resolutions (i.e., the number of agents per sector-county). Within each set of runs at a given resolution, we vary only the random seed[3] used in a pseudorandom number generator to sample the probability distributions that instantiate agents; each seed yields distinct and repeatable results. Because of the computational constraints of national dGen simulations, this analysis is limited in geographic extent to California, which is the largest DPV market in the United States.

In the next section, we present the essential mechanics of a conventional run of this model and a description of the agent instantiation procedure. We also discuss how we will assess the results of model runs at different levels of agent resolution.

2.1 The dGen Model

The dGen model simulates the adoption patterns of agents, which can be thought of as representing clusters of nonresidential and residential actors with each U.S. county. The dGen modeling approach is based on the "diffusion of innovations" framework [30, 31], which assumes new technologies are sequentially adopted by customers that vary in their innovativeness. Those who adopt earlier to pay higher prices than their mainstream counterparts who drive rapid growth once prices have declined.

2.1.1 Agent Instantiation

The dGen model instantiates agents at a fixed resolution for each county and sector (i.e., residential, commercial, and industrial). For the purposes of this report, we consider commercial and industrial agents collectively to form a nonresidential sector.

Figure 1 diagrams key elements of the agent instantiation process, in which county data sets are first used to constrain the sampling with replacement of tract and associated block attributes that are assigned to the agent. Information about the agent's location is also used to constrain the sampling of an integrated set of attributes from the U.S. Energy Information Administration's

---

[2] "Distributed Generation Market Demand," NREL, https://www.nrel.gov/analysis/dgen/.
[3] A common practice in stochastic modeling is to use a random "seed" to initialize a pseudorandom number generator. In this way, modelers can exactly reproduce prior sampling using the same seed.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Residential Energy Consumption Survey (RECS) [32] and Commercial Buildings Energy Consumption Survey (CBECS) [33] data sets, which yields an agent's annual load, building type, roof style, roof area, owner occupancy status and sampling weight. Sampling again occurs with replacement proportional to the survey weight, such that the same building would occasionally be sampled more than once within a simulation. The probability of selecting the same building multiple times at the county level grows as agent resolution increases, yet multiple selections of the same building is unlikely to happen frequently enough at any geographic level to notably impact the results.

The RECS and CBECS data sets are provided publicly at the census division level, and distributions cannot be directly isolated to a particular state or county. In this study focused on California, for example, all agents draw from the Pacific Census Division (summarized in Appendix A). Aggregation in this way diminishes the visibility of variability in building characteristics at fine geographic resolutions and inhibits the ability to make county-level comparisons (or to make comparisons among states within the same census division). Note that both RECS and CBECS use an area probability sample design to facilitate the in-person data collection. Also, because observations are clustered within selected geographic areas, they are not guaranteed to reflect the building characteristics of all geographies sampled in the study. The cumulative effect of sampling from RECS and CBECS data is that the true variability of building characteristics at the state and county levels is underestimated relative to what one would expect with a more spatially resolved data set. Even so, the inclusion of variable county-level attributes (e.g., electricity rate and solar resource) is expected to enhance modeling precision at the county level.

After RECS and CBECS building characteristics are sampled, lookup tables are created that link to derived attributes such as hourly load profiles, percentage of roof area that is suitable to development, maximum power demand, and the relative weight of the agent. Building characteristics, in conjunction with previously sampled locational attributes, are also used to constrain the selection of a set of additional lidar-based building footprint data sets [34] attributes, including tilt, azimuth, roof area suitable to DPV deployment, and flat roof status. Hourly solar capacity factor profiles for a typical year and ground cover are in turn derived using lookup tables.

Figure 1 illustrates how the random sampling procedure has a hierarchical structure in which the distribution of one characteristic may depend on the realization of another; in other words, sampling of these characteristics is based on conditional distributions. Note that Figure 1 does not include additional agent attributes that are derived using lookups at local and state geographic scales (i.e., incentives and subsidies), techno-economic calculations (i.e., the size and net present value of a system in a given year), or used-defined parameterizations (i.e., degradation rate, electricity price escalation).

4
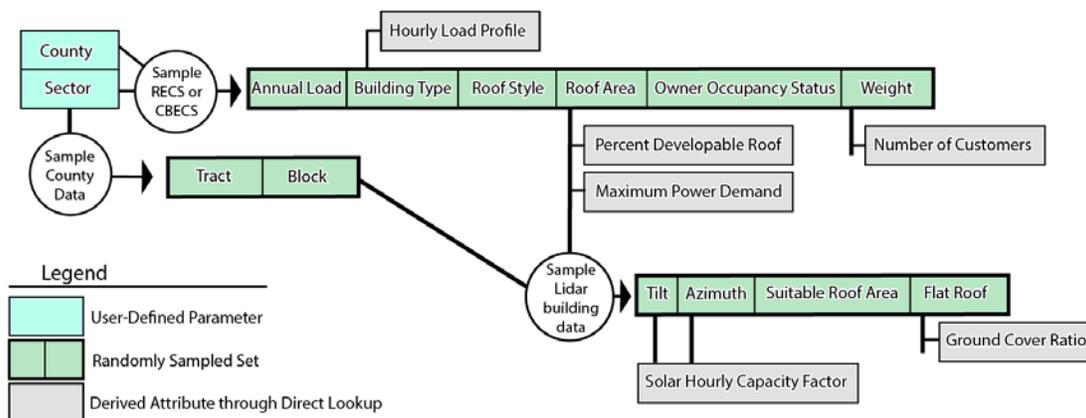
Figure 1. Instantiation workflow for key stochastically sampled agent attributes

The sampling processes outlined in Figure 1 will occur for as many agents are instantiated per county. At the lowest agent resolution of one agent per sector-county, one would expect to have low confidence in results at the county level because model results are initially based on a single draw that may not reflect the central tendency of the data set. The dGen model mitigates such variability in the sampling of weight, roof area suitable to DPV deployment, and annual load by an agent-mutation mechanism that scales these attributes in aggregate across all agents per county and sector to known totals.[4] Though the scaling ensures central tendencies are reflected, it removes heterogeneity in these attributes at the county level. Alternatively, at sufficiently high agent resolutions, the distribution in roof area suitable to DPV deployment and annual load across all agents in a county will more closely resemble the true population distribution of each sector and county. Thus, at these higher resolutions, one can be more confident that agents accurately reflect both county totals and the distributions from which they were sampled. Note that the model only constrains weight, annual load, and roof area suitable to DPV deployment to known county and sector totals; there remain attributes, including building type, load profile, and solar resource profile that vary stochastically from run to run.

2.1.2 Agent Choice Modeling

After agent instantiation and mutation, the model assesses DPV adoption decisions over time steps every two years from 2014 until 2050. During each time step, financial calculations are simulated based on exogenous technology cost declines and electricity price escalations, inclusive of state net metering policies and state and federal government subsidies. Agent decisions to adopt DPV are primarily based on the modeled financial attractiveness of DPV,

---

[4] For example, suppose there are two agents and the sampling procedure yields annual consumption values of 200 and 300. If total consumption for the county is 750, the first agent's consumption is scaled to $\frac{750}{200+300} \cdot 200 = 300$ and the second agent's consumption is scaled to 450.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

as measured by payback period, and the rate of technology adoption output from the Bass Diffusion Model.

More specifically, for each agent i in each time step t, we first derive the maximum technically-feasible system capacity for each agent, C_{it}, which is defined as the minimum of the solar capacity that would utilize the agent's entire suitable roof area, r_{it} and the solar capacity that would generate enough electricity to supply 100% of the agent's annual electrical load, l (1).

$$C_{it} = MIN(r_{it}, l) \tag{1}$$

Next, we use the C_{it} as the upper bound to determine the economically-optimal system capacity. Specifically, we search over linear increments for the system capacity, s_{it}, that maximizes the net present value of the agent's considered investment (2). Note that this implies the agent would select a null system size if the roof is unsuitable or if DPV is uneconomic.

$$s_{it} \in [0, C_{it}] \tag{2}$$

The payback period resulting from s_{it} is used to determine the fraction of customers represented through each agent that would ultimately adopt DPV (i.e., the market potential fraction, m_{it}). Recall that each agent is instantiated with a weight, w_i, thus $w_{it}m_{it}$ can be interpreted as the continuous number customers within the population subset represented by the agent that would ultimately choose to adopt DPV. The market potential fraction is derived from consumer willingness-to-pay surveys [35, 36]. The market potential fraction is updated in each year based on year-to-year changes in projected DPV payback period. Figure 2 demonstrates the modeled consumer willingness-to-pay. For a 15-year payback period, we simulate that 12% of potential residential agents and 1% of possible nonresidential agents would eventually choose to install DPV. At more-attractive payback periods, a higher fraction of agents would choose to adopt, and conversely at less attractive paybacks, fewer agents would adopt.
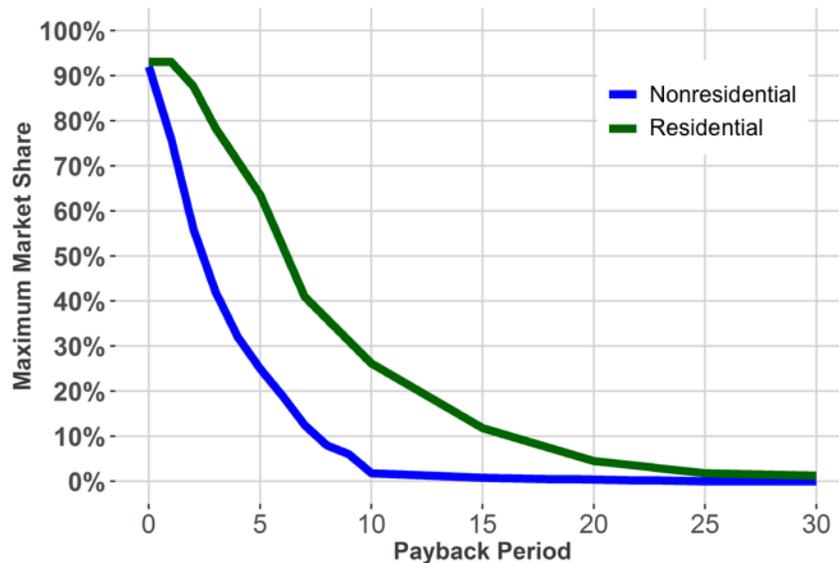
Figure 2. Market potential fraction as a function of payback period for residential and nonresidential agents

6

The market potential fraction defines the ultimate achievable level of adoption, whereas the annual number of adopters follows the Bass Model. During preliminarily modeling initiation procedures, annual new adopter counts by state and sector from 1999 to 2016 are fitted to the Bass Model using the non-linear least squares method [5, 29]. The initial year of diffusion is defined as the first year of at least five system installations, and starting conditions for the non-linear least squares regression are obtained via a genetic algorithm. The first stage is only used to estimate diffusion growth parameters (i.e., p and q). Annual adoption counts are denoted as $B(p, q, m, t)$, where t represents the time step.

Cumulative capacity at the county or state level can be calculated from the annual adoption counts following a straightforward procedure. First, let a_{it} be the cumulative number of adopters for agent i, where:

$$a_{it} = MAX(B(p, q, m_i, t) * w_{it}, a_{it-1}) \tag{3}$$

The first argument (i.e., the Bass function) returns the fraction of adoption in period t, and the second argument ensures the cumulative amount of adoption does not decrease year-on-year. Next, let x_{it} be the amount of cumulative capacity adopted by agent i, where:

$$x_{it} = a_{it} * s_{it}. \tag{4}$$

Finally, the aggregate amount of capacity adopted in county c, X_{ct}, can be calculated as:

$$X_{ct} = \sum_{i \in I_{\_c}} x_{it} \tag{5}$$

where $I_{\_c}$ represents the set of agent types in county $\_c$.

2.1.3 Model Results

We choose to focus on cumulative DPV capacity in 2050 results because this metric is commonly reported publicly and it concisely summarizes the underlying model behavior. To deconstruct primary drivers of state-level results, we also record and discuss annual new capacity, number of annual new adopters, and annual DPV system size by county and sector.

2.2 Methods for Assessing dGen Sensitivity to Agent Resolution

To diagnose the impact of varying agent resolution on projected DPV capacity, we run the model repeatedly at 1-agent, 2-agent, 10-agent, and 20-agent resolutions, using the results to calculate summary statistics and perform hypothesis testing. As summarized in Table 1, the size of each set of runs at a resolution level, denoted as n, is 100, except for the single agent level, for which $n = 1,000$. All simulations at a given agent resolution are distinct in that a unique random seed is employed for sampling purposes. Accordingly, the selection process is randomized between runs at the same resolution, and we consider the samples independent in that one run does not influence the outcome of any other run. We do allow the same seed to be used across sets of simulations by agent resolutions, which results in the sequential selection of tract-related and RECS- and CBECS-based attributes to be fixed. Accordingly, the initial distribution of tracts, utility types, annual loads (and other RECS- and CBECS-based characteristics) for a set of

agents at a low resolution will always be a subset of set of agents produced at a higher resolution given the same random seed. Randomness is still introduced in the agent instantiation process through the secondary selection of roof characteristics (i.e., tilt, azimuth, and developable roof area), and the fixing of cumulative agent roof area and developable roof area to known county totals. Thus, while we expect the duplication of seeds across agent results to yield results that are more similar than they would be if entirely different seeds were used across all agent resolutions, we still expect results to be sufficiently distinct.

For each agent resolution level, we directly calculate the mean and standard deviation of cumulative DPV capacity in 2050 for the first 100 distinct model runs. We then perform two-tailed t-tests to compare the means of the multi-agent runs (each set of runs consisting of 100 observations produced by distinct seeds) to the 1-agent resolution reference case[5] and to each other. Similarly, we perform Brown-Forsythe tests to compare the variances across different agent resolutions [37].

Table 1. Number of Model Runs for Each Level of Agent Resolution

Number of Agents per County Sector (Agent Resolution)	Number of Model Runs (n)
1	1,000
2	100
10	100
20	100

The total set of 1-agent runs is larger than the sets for the other resolutions so that we can assess whether grouped sets of dGen simulations at low resolutions yield results equivalent to higher resolution simulations. For this analysis, we compare the 1,000 1-agent runs to 100 distinct 10-agent runs, effectively normalizing the total number of agents per county-sector at 1,000 across all runs within each set. We first group the ordered set of 1,000 1-agent resolution runs into 100 bundles of 10 1-agent resolution runs, such that each bundle contains results entirely distinct from any other bundle. We then take the mean and standard deviation within each of these 100 bundles and treat this derived set as a simulated 10-agent data set. The mean and variance of the simulated 10-agent data set is then compared against the 10-agent simulations using two-tailed t-tests.

Within our analysis, we find some significant differences in mean and variance. For those instances, we unpack the model results and employ data visualization techniques to identity sources of variance in derived attributes and model inputs.

---

[5] For hypothesis testing this set consists of the first 100 observations produced by distinct seeds.

8

3 Results

In this section, we describe the results of our comparison of sets of model runs with different agent resolutions.[6]

3.1 State-Level and Sector-Level DPV Capacity by Agent Resolution

First, we compare the means and variances of the first 100 runs conducted at each agent resolution. In Figure 3, we plot state-level DPV capacity estimates over all time steps, representing mean estimates as lines and error bands representing two standard deviations from the mean. The figure illustrates that across all agent resolutions, mean estimated DPV capacity and the variance in DPV estimates increases over time until about 2040. After this time, new adoption slows and the variability among results at each agent resolution level stabilizes. Moreover, the figure reveals a key finding that both the mean and dispersion among estimates of DPV capacity across the first 100 runs is consistently greatest at the 1-agent resolution and decreases as the resolution approaches 20-agents per county. At the state level in 2050, as summarized in Table 2 (page 11), we see that the variance of 20-agent resolution runs is about one-tenth that of 1-agent runs and that the mean across all runs is reduced 6.5%. Moreover, as illustrated in Appendix B, the distribution of results for 2050 moves from a dispersed, possibly bimodal distribution to a more precise and normal distribution as agent resolution increases.
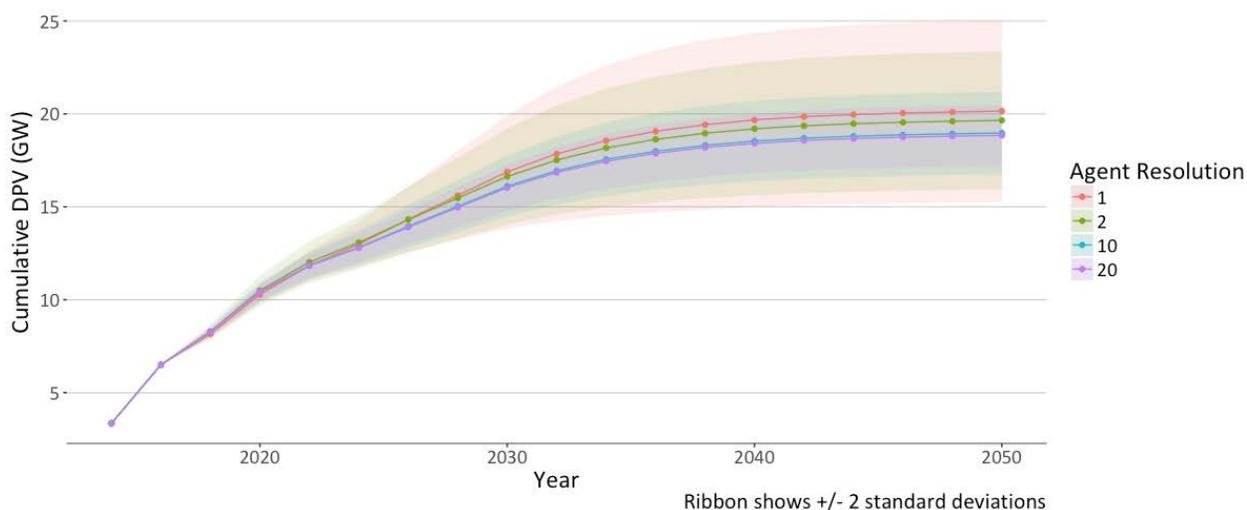
Figure 3. Estimated mean cumulative DPV capacity by year and agent resolution

We next dissect these results to understand the drivers of observed trends. We find that across all time steps, nonresidential DPV capacity correlates significantly with state-level capacity ($p < 0.01$, Pearson's r=0.97), and that this correlation is stronger than is seen in the residential sector capacity ($p < 0.01$, Pearson's r=0.94). See Appendix C for an illustration of how the nonresidential sector results closely resemble the combined state-level results in terms of the median and dispersion among estimates by year and agent resolution. Moreover, we provide

---

[6] The magnitude of estimates reported in this section are not intended to be interpreted as likely outcomes. Rather, we are concerned with the repeatability of model outputs and executing a generalizable process for identifying the source of meaningful distinctions in the central tendency of model runs at varying agent resolutions.

9

additional depth on DPV capacity estimates by sector in 2050 in Figure 4. In it, mean estimates are represented as lines and error bands represent two standard deviations from the mean. We see that mean DPV capacity decreases as agent resolution increases in both the nonresidential and residential sectors, with the nonresidential sector decreasing from a DPV estimate of and a standard deviation of 2.45 GW at a 1-agent resolution, to a DPV estimate of 9.8 GW and a standard deviation 0.8 GW at a 20-agent resolution. In the residential sector, we similarly see a capacity decrease from 9.4 GW to 9.0 GW, and a standard deviation reduction from 0.2 GW to less than 100 MW. Mean and variance decreases are most pronounced in the nonresidential sector, where mean DPV estimates at the 20-agent resolution are about 9% lower than at the 1-agent resolution; in the residential sector, we see less than a 4% decline across resolutions. We clearly see that that variance is much lower in the nonresidential sector and find that the variances between sectors are in fact significantly different ($p < 0.01$).
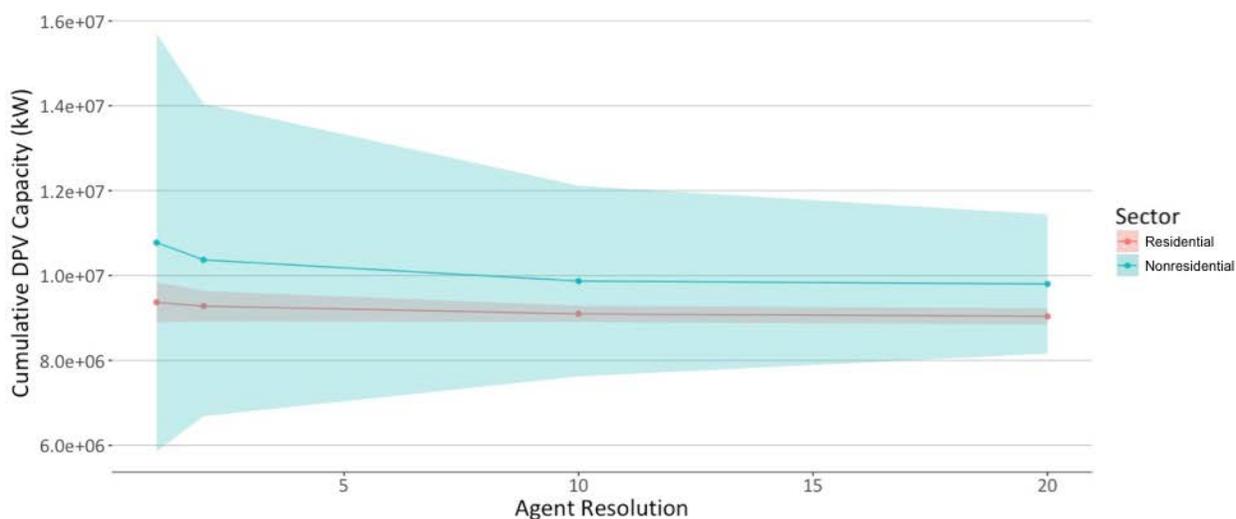
Figure 4. Estimated cumulative DPV capacity in 2050 by sector and agent resolution

In addition to the overall decreases in mean and variance that we observe in the extreme agent resolution cases (1-agent compared to 20-agent), we also see consistent decreases for the intermediate resolutions. Moving from a 1-agent to 2-agent resolution, the cumulative DPV capacity across all sectors and simulated runs in 2050 estimate decreases by 2.5% to 19.7 GW. Though these aggregate state means are not statistically significantly different in aggregate, the residential mean does decrease by 1%, which is significant at the 5% level (i.e., $p < 0.05$). This result suggests the residential sector is more sensitive to resolution changes in this spectrum than the nonresidential sector and that this sector plays a less dominant role in state-level trends. Regarding variance in state-level 2-agent resolution results in 2050 compared to the 1-agent runs, variance decreases by more than 40% as the standard deviation decreases to 9.4% of the cumulative DPV mean; however, the significance of this difference is relatively weak ($p < 0.10$). Within each sector, we see declines in variance with weak significance.

We also find that 10-agent runs tend towards a mean DPV capacity that is lower than 1-agent runs, while also demonstrating higher precision. For a 10-agent per county resolution, the estimated cumulative DPV capacity across all sectors and simulated runs in 2050 decreases by 6% compared to the 1-agent resolution runs ($p < 0.01$). The DPV estimate is 8.5% lower in

nonresidential sector ($p < 0.01$) and only 3% lower in the residential sector ($p < 0.01$). The overall state-level variance decreases to one-fifth that of the 1-agents runs ($p < 0.01$) and standard deviation declines from 2.4 GW to 1.1GW. We find similar results within each sector as well; at a 10-agent resolution, the residential and nonresidential sector variances are less than one fifth that of the 1-agent runs ($p < 0.01$), and standard deviations respectively decrease to 1% and 11% the cumulative sector mean.

When comparing 10-agent to 2-agent resolution runs, we find again reduced mean DPV estimates and variance. Cumulative DPV estimates decrease both in the residential (-2.0%, $p < 0.01$) and nonresidential (-4.8%, $p < 0.05$) sectors. At the state level, variance is reduced by approximately two-thirds ($p < 0.5$). Similar reductions are seen at the sector level, with variance being most reduced in the residential sector.

As the resolution increases to 20-agents per county, we notice diminishing mean and variance changes. The cumulative DPV capacity estimates for the state and for the nonresidential sector are not significantly different than at 10-agents per county. Moreover, though there is a statistically significant decline in cumulative DPV capacity in the residential sector in 2050 ($p < 0.01$), it is at a rate of 0.7%. However, variance of the 20-agent resolution runs in the nonresidential sector (-46%, $p < 0.01$) is half that of the 10-agent runs. Residential variance and variance at the state level do not significantly decline between 10-agent and 20-agent resolutions.

Table 2. Summary of Two-Way Comparisons across Agent Resolution Levels of Statewide Cumulative DPV Capacity in 2050

Base Resolution	Comparison Resolution	Percent Decrease Mean	Ratio of Variance, Comparison : Base
1	2	2.4	0.58[a]
1	10	5.9[a]	0.21a
1	20	6.5[a]	0.21[a]
2	10	3.5%[a]	0.35[a]
10	20	-0.7%	0.55[a]

[a] Denotes $p < 0.01$

3.2 Equivalence of DPV Capacity by Agent Resolution

Having found distinctions in DPV capacity by agent resolution, we briefly address the question of whether bundles of 10 1-agent simulation results are equivalent to 10-agent resolution runs on average at the state level. We group 1,000 1-agent results into a simulated 10-agent data set (n=100), with each bundle of observations being the mean of 10 distinct output results. As would be expected from previous results, we find that cumulative DPV capacities in the simulated data sets do not converge on the same mean as the 100 10-agent runs ($p < 0.01$). However, variance distinctions are perhaps less intuitive to anticipate, and here we find they are dissimilar as well. Empirically, the simulated data set variance is less than 3% that of the 10-agent runs ($p < 0.01$).

3.3 State-Level and Sector-Level Trends

Having discovered meaningful differences in mean estimated DPV capacity at varying agent resolutions, we next seek to identify the mechanisms in agent instantiation and subsequent internal dGen processing procedures accountable for these divergent outcomes. We begin with data exploration and visualization that are guided by model intuition.

Logically speaking, amounts of new installed DPV capacity (see Appendix D) are subject to exogenous economic conditions and are proportional to the number of new customers and the size of their system. Across all years and agent-resolution simulations, univariate linear regression suggests that observed new DPV capacity better aligns with the number of new adopters[7] than it does with the weighted averaged installed size of a DPV system[8]. Similarly, at the sector level, we see the number of new adopters in each sector explaining a majority of that sector's new DPV adoption.[9]
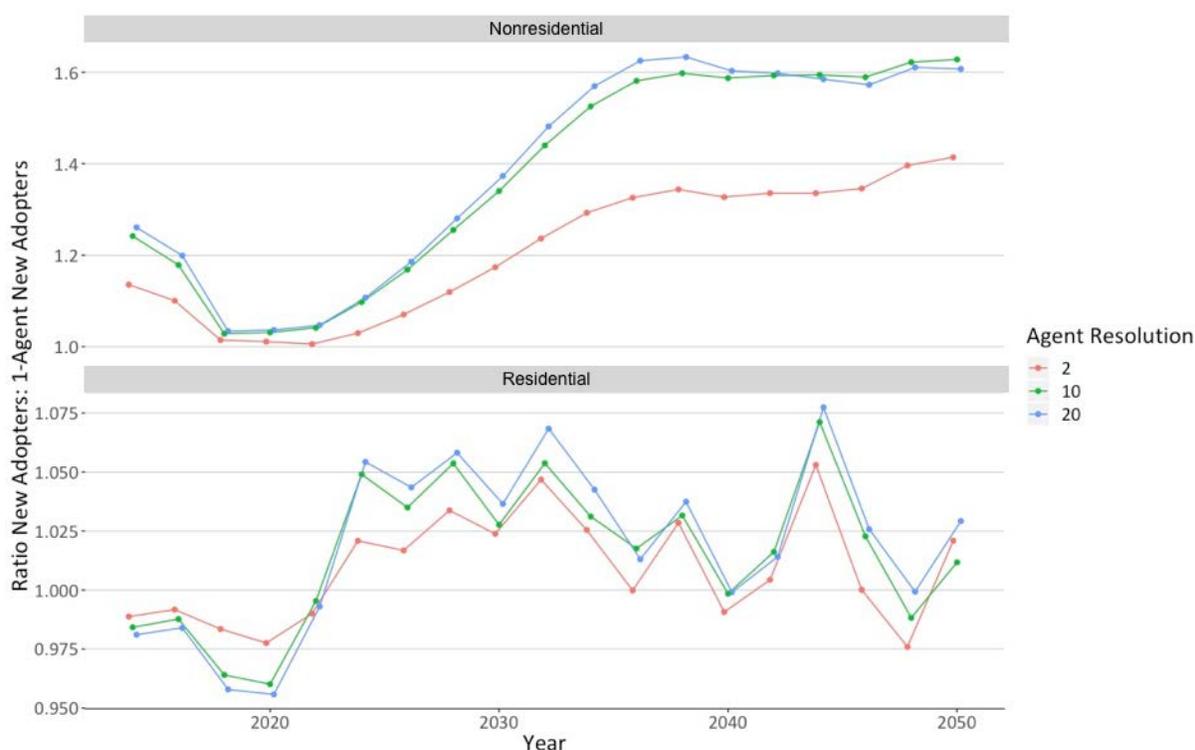
Figure 5. Ratio of number of new adopters to 1-agent new adopters by year, sector and agent resolution

---

[7] Described by C_s = .00002453 + 1.724*x, where C_s is new DPV capacity by year and agent resolution, x is total number of new adopters for that year and agent resolution; adjusted R-squared is 0.8108.

[8] Described by C_s = 2055593 - 287418*x, where C_s is new capacity by year and agent resolution, x is weighted average DPV system size for that year and agent resolution; adjusted R-squared is 0.13.

[9] Residential DPV Capacity is described by C_r = -.00167 + 1.156*x_r, where C_r is new residential capacity by year and agent resolution, x_r is number of new residential adopters for that year and agent resolution. Adjusted R-squared is 0.986.
Nonresidential DPV Capacity is described by C_n = 19710 + 36.18*x_n, where C_n is new nonresidential capacity by year and agent resolution, x_n is number of new nonresidential adopters for that year and agent resolution. Adjusted R-squared is 0.744.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Though the number of adopters may explain the magnitude of new DPV adoption well, as revealed in Figure 5, this metric does not consistently decrease as agent resolution increases. Instead, across all years in the nonresidential sector, we observe a pattern of high agent resolution runs having increasingly greater numbers of new adopters on average than 1-agent resolution runs. In the residential sector, consistent trends in new adoption counts relative to the 1-agent mean are not evident across all years. Though these phenomena are interesting on their own, we will continue to search for variables that explain the sector and state-level results.

We next examine the mean size of installed DPV systems by agent resolution in the residential and nonresidential sectors (Figure 6). Here we find within each year after 2016 that the average size of a customer's system consistently decreases with increasing agent resolution in both sectors. Thus, we conclude that DPV size is counteracting the trends in the number of new adopters to drive the new DPV installation outcomes that we see by agent resolution at the state (Figures 3) and sector (Figure 4) levels.
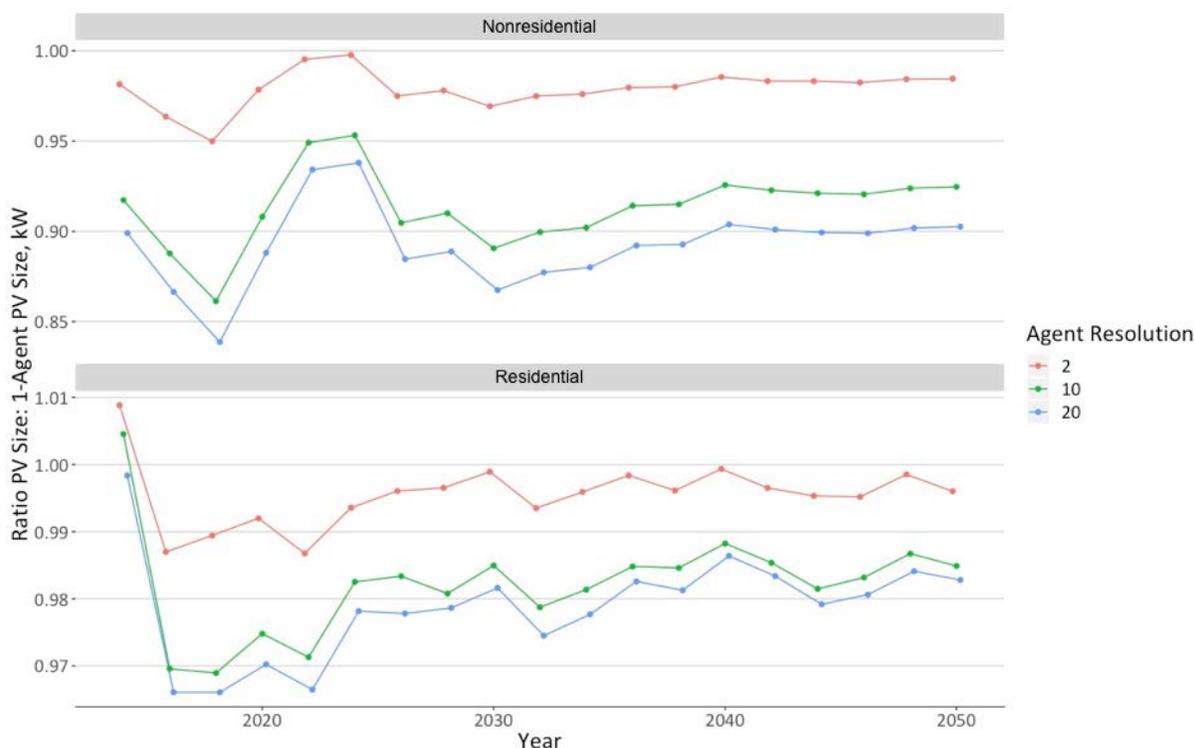
Figure 6. Ratio of mean PV size to 1-agent mean PV size by year, sector and agent resolution

We attribute the decrease in the weighted average size of a DPV installation at the sector level to the system-sizing mechanism in dGen that scales maximum system size to the lesser of that which the suitable roof size or annual load will support. Recalling that counties have fixed total county load and suitable roof areas, either metric independently will produce a suitable average supported DPV size at the county level. Moreover, independently the distribution of supported system sizes produced by either metric across all agents within the county will reflect the underlying building characteristic distributions from which agents are instantiated.

13

At the same time, however, supported system sizes derived from suitable roof size or annual load are not ensured to settle on the same average at the county level. More importantly, in selecting the minimum between the two possible supported sizes on an agent by agent basis within counties, we ensure the average selected system size at the county level will never be more than the lesser of the two alternatives. Furthermore, within a county, if supported DPV sizes are not consistently greater on an agent by agent basis according to either suitable roof size or annual load, the average selected DPV size will be less than what either metric alone would produce (see Appendix E for an illustration of these points). Given that suitable roof area and annual load are not ensured to be positively correlated at the agent level, it is increasingly likely that as agent resolution increases, one set of possible systems sizes will not be consistently greater, thus diminishing the average DPV system size over many iterations (see Appendix E).

3.4 County-Level DPV Capacity by Agent Resolution

Next, we examine DPV capacity estimates at the geographic extent at which agents are instantiated: the county.[10] Provided that key building characteristics are sampled from RECS and CBECS data sets aggregated from select counties at the Census division level, we recognize that results will be inherently unreliable at the county level. An analysis of county-level results does provide a baseline understanding of variability, which provides context for the data and model enhancements that would facilitate interpretation at this granularity.

We find that as agent resolution increases, the state-level trend of decreasing mean DPV capacity does not consistently hold. Figure 7 (next page) illustrates, for example, that while nonresidential county-level DPV estimates generally decrease with agent resolution, nearly one in five counties located mostly in southeastern California demonstrate increasing DPV capacity with increasing agent resolution. Moreover, we find that the magnitude in DPV fluctuations is amplified at the county level; in the nonresidential sector, we observe as much as a 40% decrease or 230% increase in estimated DPV capacity between the 1-agent and 20-agent resolutions. These same types of divergences are evident in the residential sector results as well, though in half as many counties, the fluctuations are less dramatic. For example, the mean among 20-agent resolution runs does not differ more by than 25% compared to the 1-agent resolution mean in any county. We expect that in these counties, trends in numbers of new adopters are outweighing the previously observed mechanisms reducing average system sizes.

Variance at the county level is higher than at the state or sector levels. In the nonresidential sector, the standard deviation ranges from 3% to 34% of the county sector mean (22% on average), while in the residential sector it varies more closely between 2% and 8% of the county sector mean (4% on average). We generally see that variance declines with increasing agent resolution; however, in more than half the counties in the residential sector and in one-fifth of the counties in the nonresidential sector, we observe an increase in variance between the 1-agent and 2-agent resolutions. At a 20-agent resolution, the proportion of counties that sustain this increase in variance is only 5% of counties in the residential sector and 10% of counties in the nonresidential sector.

---

[10] California has 58 counties with 1-agent mean DPV capacity estimates in 2050 ranging from 0.1 MW to 2.3 GW.
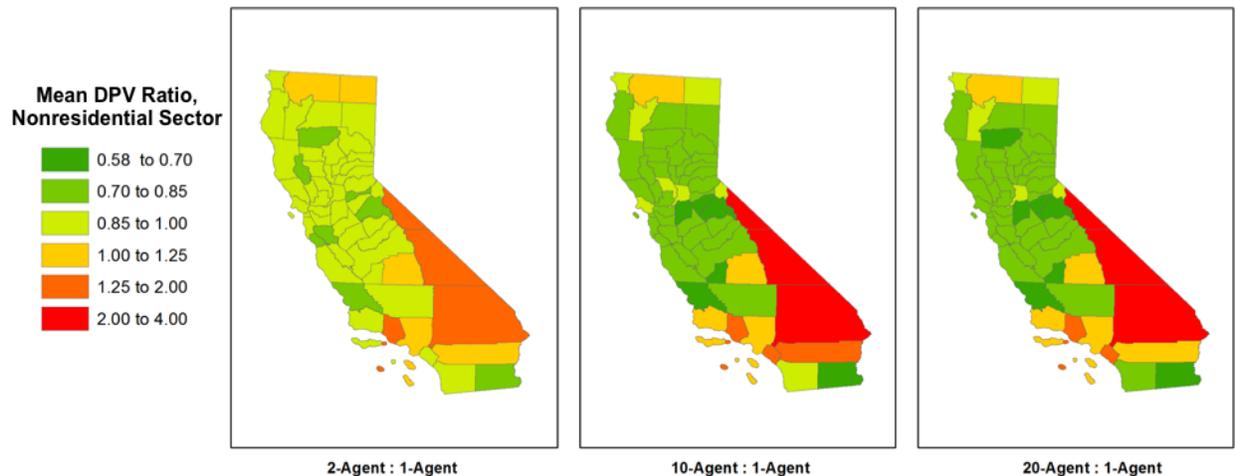
14

Figure 7. Ratio mean nonresidential county-level DPV capacity estimates in reference to 1-agent resolution runs

We employ a least-squares linear regression to correlate county-level attributes to standard deviation trends. In the nonresidential sector, we find that by aggregating agents at the county level and then averaging attributes across the first 100 simulations at each resolution, we can account for 60% (adjusted R-squared = 0.607) of the observed variability in standard deviation by county. Key explanatory variables include maximum observed capacity factor, the cross product of agent resolution, suitable roof area, annual load, as well the number of developable customers (all coefficients are significant at $p < 0.001$, except suitable roof area and its cross product with agent resolution). In isolation, the average range of observed capacity factors and the average suitable roof area of a county each account for up to 30% of observed trends in county standard deviations relative to the mean ($p < 0.001$, adjusted R-squared values of 0.28 and 0.27 respectively). The correlation with the number of nonresidential developable customers in a county is significant at $p < 0.01$, though it accounts for relatively little of the observed variance (adjusted R-squared = 0.05). In the residential sector, the variance in standard deviation relative to the mean is not significantly explained by the same set of attributes, except for agent resolution.

3.5 Precision by Agent Resolution

As discussed previously in Section 3.1, nonresidential and residential results are seen to become more precise as agent resolution increases at the state level. We also noted that variance at the state level mostly comes from the nonresidential sector, and that this sector would require a substantially higher resolution than we modeled to achieve variance that is similar to that seen in the residential sector. From empirical results, we fit a log function to the nonresidential and residential state-level results and then plot trend lines in Figure 8. Based on the projections illustrated in this graphic, in the nonresidential sector (intercept – 0.22007, log=-0.04628), we find that at a 76-agent resolution is needed to obtain less than 2% standard deviation, and a 94-agent resolution would be expected to bring deviation within less than 1%. Similarly, in the more inherently precise residential sector (intercept – 0.024129, log=-0.005094), a 17-agent resolution is expected to be sufficient to bring the standard deviation to within less than 1%. Given that we empirically find a standard deviation slightly greater than 1% at a 20-agent resolution, however, we interpret these estimates as approximate indicators of suitable agent resolution ranges.
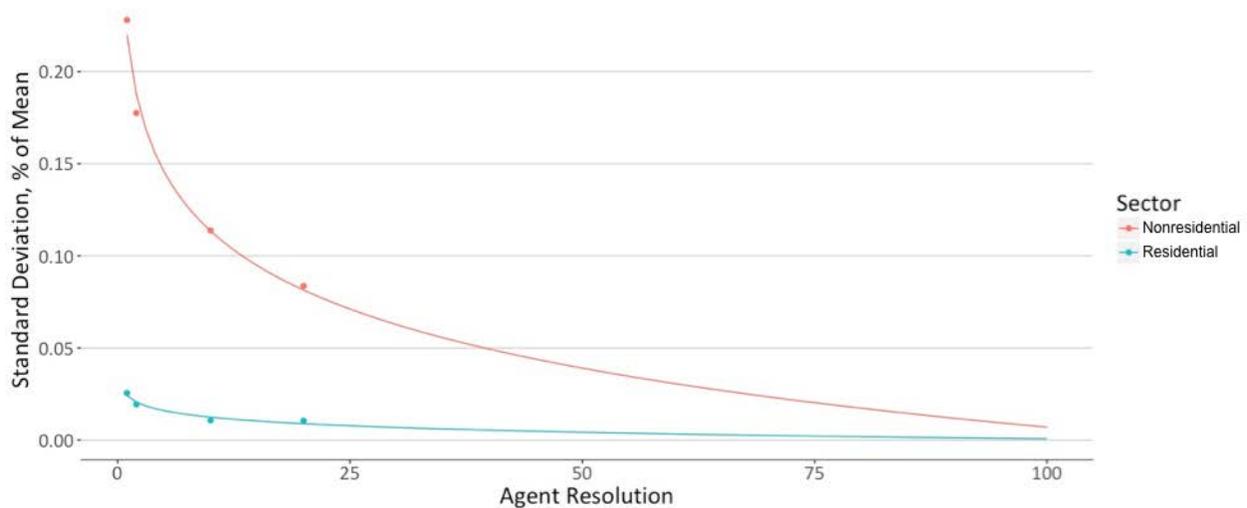
Figure 8. Projected state-level standard deviation reductions with increasing agent resolution by sector

3.6 Precision by Number of Runs

We next employing bootstrap methodologies to examine how the precision of the first 100 uniquely seeded runs fluctuates. For each agent resolution, at each number of runs between 2 and 100, we sample a set of distinct 2050 DPV estimates sized to that number of run 100 times. For each sample, we measure the difference in its standard deviation and that of the first 100 runs at the same agent resolution. As plotted results reveal in Figure 9 (next page), the mean difference in standard deviation decreases asymptotically as the number of runs approaches that 100 runs at each agent resolution. Also, in general we see that as agent resolution increases, the rate at which intermediate variance approaches that of 100 runs slows.

It is important to remember when interpreting Figure 9 that the actual magnitude of standard deviation at 100 runs in relation to the mean of those run varies by agent. For example, consider running dGen at 1-agent and 20-agent resolutions, twenty times each. Figure 9 suggests we would observe approximately 10% and 14% differences from the 100-agent standard deviation for the 1-agent and 20-agent resolution runs respectively. Recalling that the standard deviation at the state level is about 4% of the mean a 20-agent resolution at 100 runs, we would accordingly not expect the standard deviation relative to the mean to fluctuate more than 0.6%. However, at a 1-agent resolution where standard deviation is 12% of the mean at 100 runs, we would expect the standard deviation as a percentage of mean to fluctuate by twice as much. More than two times as many runs at the 1-agent resolution would be needed to be match the magnitude of fluctuation seen at a 20-agent resolution. However, it is worth noting these standard deviation distinctions are not expected to dramatically impact the interpretation of results. This example best serves to highlight that the inherent variance of an agent resolution may be more influential in achieving an acceptable level of variance than the number of times at which that resolution is run.
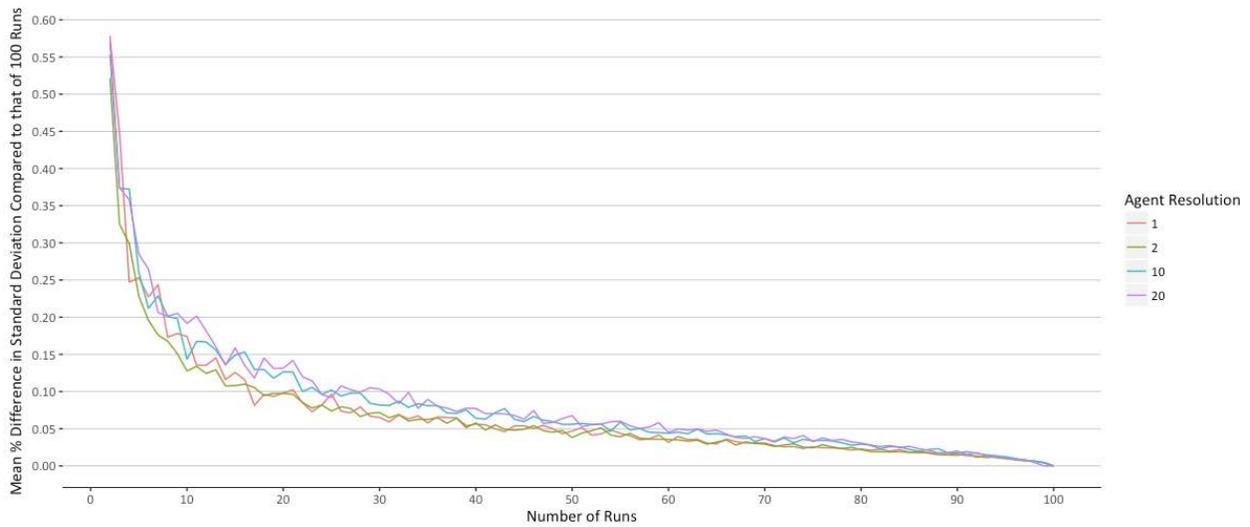
Figure 9. Mean percent difference in standard deviation compared to that of the first 100 runs by agent resolution

3.7 Average Run Time by Agent Resolution

In this section, we report the average run time across all simulations at 1-agent (n=1000), 2-agent (n=100), 10-agent (n=100), and 20-agent (n=100) resolutions. As illustrated in Figure 10, we moreover fit a simple linear regression (adjusted R=0.90) to observed means and find that from an approximate minimum solve time of 18 minutes, each agent adds approximately three minutes to the total run time.
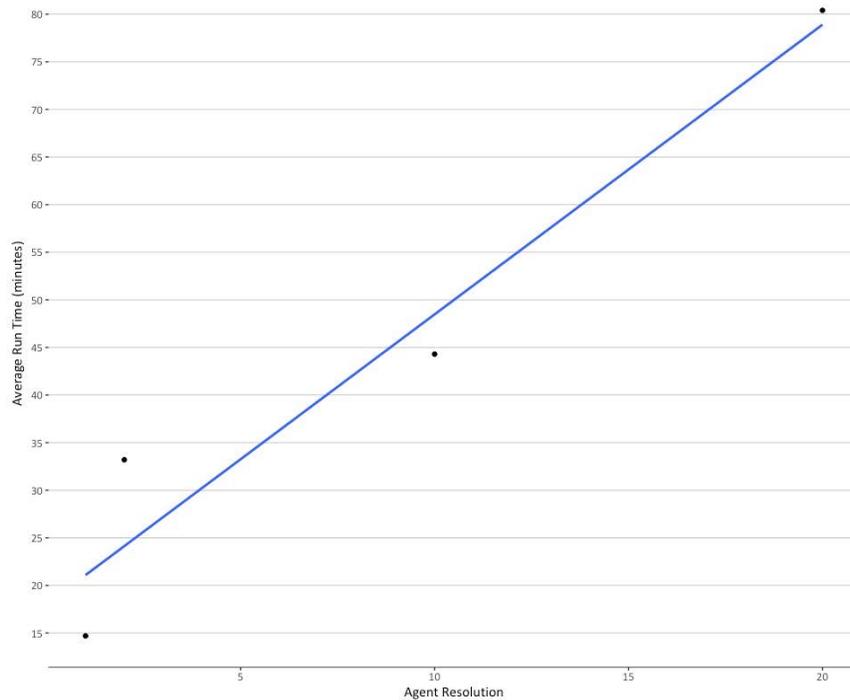
Figure 10. Average run time in minutes by agent resolution

17

4 Discussion

In this report, we explore trade-offs of top-down and bottom-up methods in their precision and computational burden using NREL's dGen model, an agent-based model of residential and nonresidential distributed PV adoption. Running the model for a single state with fixed macroeconomic assumptions, we vary both the number of agents instantiated and the number of unique simulations. Notably, we introduce sources of uncertainty in cumulative DPV adoption results during agent instantiation as we stochastically sample agent attributes from localized building characteristic and solar resource distributions. The composition of these distributions and the number of times that we sample from them have implications for the tendency of observed results, and we find three key trends that guide what can be expected of results. First, precision of the model projections, as measured by the variance of the unique simulations increased as a function of the number of agents sampled, though with diminishing returns to scale. Second, the mean or expected value of projections decreased as the number of agents instantiated increased. Finally, the variance of model projections differed by sector, where variance was substantially larger in the commercial sector, largely because of differences in parameterization by sector.

We explain the results both as a function of how the dGen model is constructed as well as known statistical phenomena. The first result, that variance decreases as sampling rate increased, is self-explanatory through the law of large of numbers—namely that we are more fully representing the underlying distributions instantiating agents. However, the second and third results require more consideration. For the second result, we demonstrate that projected capacity decreased as agent resolution increased due to a corresponding decrease in the size of PV systems agents selected to install. Crucially, the systems are smaller due to the system-sizing mechanisms tied to independent suitable roof area and annual load distributions that biases smaller system sizes at larger agent resolutions. Model improvement mechanisms that serve to more directly correlate these distributions at agent instantiation will help more sensibly reflect system size at the agent level.

We explain the larger variance in the nonresidential sector as an outcome of the greater variance in building uses and load shapes as compared to the residential sector. Note that nonresidential building use cases span from small commercial offices to large industrial complexes. This diversity of building uses and their resulting load shapes affects the optimal size of the PV system for each agent and therefore requires more agents to accurately represent does the residential sector. At a minimum, a sensible approach to efficiently improving dGen precision could be to parameterize agent resolution with higher numbers of nonresidential agents than residential agents. More-advanced model enhancements in the nonresidential sector, including further agent stratification (i.e., building classification, tariff structure, and land use type) and distinct financial and adoption models by customer user case, would also help improve both the precision and accuracy of results.

A perennial challenge in systems modeling is deciding the optimal trade-off between model resolution and computational feasibility. For example, at a 20-agent resolution, the coefficient of variance is 1% in the residential sector, and eight times higher in the nonresidential sector. Improving nonresidential variance to 1% or less in both sectors would require nearly 100 agents per county and subsequently, roughly 22 days of computation. Because reducing the number of runs to as low as 20 iterations is found not to substantially impact the interpretation of results relative to 100 iterations; however, that computational time could be scaled down to five days with modest precision losses.

Ultimately, more resolved models offer greater precision and more spatially-granular results, yet they are more burdensome in parameterization, sensitivities assessment, and computation resources. Our results give insight to the prospective modeler of these trade-offs, where model run time increased with the agent resolution and precision decreased as a logarithmic function. For example, our results suggest that sets of 100-agent resolution runs would demonstrate high precision in cumulative DPV capacity estimates such that the magnitude of observed standard deviations would be expected to be within 5% of the mean DPV capacity of the set. At the same time, a single run at this resolution would be expected to take more than five hours and compiling a set of 20 runs alone could take more than four days. Considering that these metrics are relevant to analyses bounded to the state of California, national results would be expected to increase computing time and resource requirements considerably. Research exploring multi-criteria optimization in terms of precision, run time, and computing capacity would help improve modeling efficiency.

Furthermore, note that both the mean and variance changed as a function of agent resolution. A priori, any single 20-agent simulation is closer to its population mean than any single 1-agent simulation, yet we would not estimate the population mean without multiple simulations. This implies that an efficient frontier exists for determining the minimum run time necessary for any given precision level.

Tradeoffs of precision and computational burden have practical implications for analysis of dGen results. For the purposes of policy analysis, a high level of precision is a prerequisite for simulating potentially subtle effects. Also, customer adoption models are increasingly expected to inform distribution system planning, for which even county-level results are too coarse. Uniformly increasing agent-resolution across all counties is a promising pathway toward higher precision at higher spatial granularity. This broad approach is subject to computational constraints, and scaling agent resolution in relation to county standard deviation or weighting county resolution by relevant county-level attributes could help achieve computational efficiencies. Resolution parameterization by county, however, requires further research to understand (1) possible biases either sampling approach would introduce to counties and (2) acceptable sampling rates for interpreting county-level results.

Our analysis also highlights that understanding model parameterization is significant to the interpretation of results when evaluating potential policy or technology decisions—in particular, one must be assured that differences in modelled outcomes are actually due to policy differences, and not sampling error. Parameterization of advanced models can be complex and interdependent such that we posit that different macroeconomic parameters (e.g., technology costs and fuel prices) will require different sampling resolutions to achieve comparable precision to results detailed in this report. Furthermore, having noted that localized building characteristics and solar resource distributions play key roles in model results, it is also important to have more robust methodologies for gauging the precision limitations at various scales based on the quality and resolution of input data. We did not explore interstate differences in precision; for instance, states with different market maturities might behave differently. Furthermore, the results are in some sense determined by arbitrary modeling relationships. For instance, the choice of 16 different nonresidential load shapes versus a single one. As new agent-level attributes are added (e.g., income or environmental concern), they potentially reduce precision by requiring more sampling to represent the fully-convolved distribution.

19

5 Conclusions

Our empirical analysis of the dGen model provides insights that are relevant to the quantification of stochastic model uncertainty. Through empirical observations at varying agent resolutions, we find that aggregate DPV capacities projected by dGen are more precise at a 10-agent resolution compared to at a 1-agent resolution. We also find results are also on average 7% lower at this higher resolution. Moreover, estimated DPV capacities are up to 10% lower at a 10-agent resolution compared to a 1-agent resolution the nonresidential sector, and this sector accounts for the majority of state-level variation. A resolution on the order of 100 agents per county would be needed to bring nonresidential variance in line with residential variance at a 2-agent resolution. We attribute precision gains at a 10-agent resolution to the Law of Large Numbers and mean DPV capacity reductions to a system-sizing scheme that biases lower cumulative installed capacity as agent resolution increases when roof size and annual load consumption attributes are not positively correlated. We propose that further research explore increasing agent resolution in the nonresidential sector and more clearly quantify sources of variability at the county level such that resolution can by dynamically increased where county-level results are least precise.

6 References

1. GTM Research U.S. Solar Market Insight, Full Report Q2 2018.

2. St. John J. (2018) Distributed Energy Poised for 'Explosive Growth' on the US Grid. *Greentech Media*. Available at: https://www.greentechmedia.com/articles/read/distributed-energy-poised-for-explosive-growth-on-the-us-grid#gs.r8sQwN4.

3. Graziano M., Gillingham K. (2015) Spatial Patterns of Solar Photovoltaic System Adoption: The Influence of Neighbors and the Built Environment. *J Econ Geogr* 15(4):815–839.

4. Mills A., et al. (2016) *Planning for a Distributed Disruption: Innovative Practivces for Incorporating Distributed Solar into Utility Planning.* (Lawrence Berkley National Laboratory, Berkely, CA).

5. Dong C., Sigrin B., Brinkman G. (2017) Forecasting Residential Solar Photovoltaic Deployment in California. *Technol Forecast Soc Change* 117:251–265.

6. Davidson C., Drury E., Lopez A., Elmore R., Margolis R. Modeling Photovoltaic Diffusion: An Analysis of Geopspatial Datasets. *Env Res Lett* 9(7). doi:074009.

7. Bernards R., Morren J., Slootweg H. (2018) Development and Implementation of Statistical Models for Estimating Diversified Adoption of Energy Transition Technologies. *IEEE Trans Sustain Energy* PP(99):1.

8. Dharshing S. (2017) Household dynamics of Technology Adoption: A Spatial Econometric Analysis of Residential Solar Photovoltaic (PV) Systems in Germany. *Energy Res Soc Sci* 23:113–124.

9. Kwan C. (2012) Influence of Local Environmental, Social, Economic and Political Variables on the Spatial Distribution of Residential Solar PV Arrays Across the United States. *Energy Policy* 47:332–344.

10. Williams E., Carvalho R., Ronneberg M. (2017) International Model for Diffusion of Residential Solar Power.

11. Agarwal A., Cai D., Shah S., Chandy M., Sherick R. (2015) A Model for Residential Adoption of Photovoltaic Systems. *2015 IEEE Power & Energy Society General Meeting* (Denver, CO), pp 1–5.

12. Wang W., Yu N., Johnson R. (2017) A Model for Commercial Adoption of Photovoltaic Systems in California. *J Renew Sustain Energy* 9(025904). Available at: https://doi.org/10.1063/1.4979899.

13. Guidolin M., Mortarino C. (2010) Cross-Country Diffusion of Photovoltaic Systems: Modelling Choices and Forecasts for National Adoption Patterns. *Technol Forecast Soc Change* 77(2):279–296.

14. California Energy Commission (CEC) (2015) *California Energy Demand 2018-2028, Preliminary Electricity Forecast* Available at: https://efiling.energy.ca.gov/getdocument.aspx?tn=220615.

15. EIA (U.S. Energy Information Administration) (2013) *Residential Demand Module of the National Energy Modeling System: Model Documentation* Available at: http://www.eia.gov/forecasts/aeo/nems/documentation/ residential/pdf/m067(2013).pdf.

16. Rai V., Robinson S. (2015) Agent-based Modeling of Energy Technology Adoption: Empirical Integration of Social, Behavioral, Economic, and Environmental Factors. *Environ Model Softw* 70(August):163–177.

17. Alyousef A., Adepetu A., de Meer H. (2017) Analysis and Model-based Predictions of Solar PV and Battery adoption in Germany: An Agent-based Approach. *Comput Sci - Res Dev* 32(211). Available at: https://doi.org/10.1007/s00450-016-0304-9.

18. Mittal A., Huang W., Krejci C. (2017) Consumer-adoption Modeling of Distributed Solar Using an Agent-based Approach. *Comput Soc Sci*. Available at: http://works.bepress.com/anuj-mittal/11/.

19. Denholm P., Drury E., Margolis R. (2009) *The Solar Deployment Systems (SolarDS) Model: Documentation and Sample Results* (National Renewable Energy Laboratory, Golden, CO).

20. Darghouth N., Wiser R., Barbose G., Mills A. (2016) Net Metering and Market Feedback Loops: Exploring the Impact of Retail Rate Design on Distributed PV Deployment. *Appl Energy* 162:713–722.

21. Adepetu A., Keshav S. (2016) Understanding Solar PV and Battery Adoption in Ontario: An Agent-based Approach. *In Proceedings of the Seventh International Conference on Future Energy Systems (e-Energy '16)* (ACM, New York, NY, USA). Available at: https://doi.org/10.1145/2934328.2934333.

22. Palmer J., Sorda G., Madlener R. (2013) *Modeling the Diffusion of Residential Photovoltaic Systems in Italy: An agent-based simulation*.

23. Murakami T. (2014) Agent-based Simulations of the Influence of Social Policy and Neighboring Communication on the Adoption of Grid-connected Photovoltaics. *Energy Convers Manag* 80:158–164.

24. Zhao J., Mazhari E., Celik N., Son Y. (2011) Hybrid Agent-based Simulation for Policy Evaluation of Solar Power Generation Systems. *Simul Model Pract Theory* 19(10):2189–2205.

25. Zhang H., Vorobeychik Y., Letchford J., Lakkaraju K. (2015) Data-driven Agent-based Modeling, with Application to Rooftop Solar Adoption. *In Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pp 513–521.

26. Ajzen I. (1991) The Theory of Planned Behavior. *Organ Behav Hum Decis Process* 50(2):179–211.

27. Sigrin B., Gleason M., Preus R., Baring-Gould I., Margolis R. (2016) *The Distributed Generation Market Demand Model (dGen): Documentation.* (National Renewable Energy Laboratory, Golden, CO) Available at: http://www.nrel.gov/docs/fy16osti/65231.pdf.

28. Drury E., Jenkin T., Jordan D., Margolis R. (2014) Photovoltaic Investment Risk and Uncertainty for Residential Customers. *IEEE J Photovolt* 4(1).

29. Srinivasan V., Mason C. (1986) Technical Note—Nonlinear Least Squares Estimation of New Product Diffusion Models. *Mark Sci* 5(2):169–178.

30. Bass F.M. (1969) A New Product Growth for Model Consumer Durables. *Manag Sci* 18:215–227.

31. Rogers E. (2003) *The Diffusion of Innovation, 5th edition* (Free Press, New York).

32. EIA (U.S. Energy Information Administration) (2014) 2009 Residential Energy Consumption Survey Data. Available at: http://www.eia.gov/consumption/residential/data/2009/ index.cfm?view=microdata (Accessed March 2, 2014).

33. EIA (U.S. Energy Information Administration) (2008) 2003 Commercial Building Energy Consumption Survey (CBECS) Public Use Microdata. Available at: http://www.eia.gov/consumption/commercial/data/2003/index.cfm?view=microdata (Accessed March 2, 2014).

34. Gagnon P., Margolis R., Melius J., Phillips C., Elmore R. (2015) *Rooftop Photovoltaic Technical Potential in the United States: A Detailed Assessment* (National Renewable Energy Laboratory, Golden, CO).

35. Sigrin B., Drury E. (2014) Diffusion into New Markets: Economic Returns Required by Households to Adopt Rooftop Photovoltaics. *AAAI Energy Market Prediction Symposium*.

36. Paidipati J., Sawyer H., Kurrasch A. (2008) Rooftop Photovoltaics Market Penetration Scenarios. *Navig Consult Inc NREL Febr*.

37. Brown M., Forsythe A. (1974) Robust Tests for the Equality of Variances. *J Am Stat Assoc* 69(346). doi:10.1080/01621459.1974.10482955.

Appendix A. RECS and CBECS Summary for the Pacific Census Division

The following Table A-1 summarizes the key descriptive statistics of the Pacific Census Division subset of the RECS and CBECS data sets used in this analysis. From this subset, dGen samples annual load, building type, roof style, roof area, owner occupancy status and weight.

**Table A-1. Summary of RECS and CBECS in the Pacific Census Division**

Building Type	Sector[a]	Cumulative Weight	Mean Roof Area (sq. ft.)	Std. Dev. Roof Area (sq. ft.)	Mean Annual Load (kWh)[b]	Std. Dev. Annual Load (kWh)
Full Service Restaurant	Non-Res.	52,200	5,200	4,600	197,000	185,400
Hospital	Non-Res.	27,700	9,600	19,300	2,188,800	528,600
Large Hotel	Non-Res.	5,400	32,200	35,700	2,628,200	2,109,900
Large Office	Non-Res.	14,600	24,500	16,800	2,831,700	2,318,900
Medium Office	Non-Res.	103,900	8,600	8,900	803,300	225,000
Midrise Apartment	Non-Res.	6,000	19,800	14,500	320,600	404,900
Out Patient	Non-Res.	34,200	5,900	8,400	341,100	139,700
Primary School	Non-Res.	59,300	8,100	17,000	436,600	124,000
Quick Service Restaurant	Non-Res.	16,500	4,000	2,500	208,700	308,400
Reference Home	Residential	9,359,700	1,800	900	7,500	10,400
Secondary School	Non-Res.	134,800	7,600	14,800	249,700	107,800
Small Hotel	Non-Res.	16,900	10,900	19,400	203,200	99,100
Small Office	Non-Res.	329,200	3,700	4,300	153,800	63,600
Stand Alone Retail	Non-Res.	143,000	11,500	21,700	501,200	157,700
Strip Mall	Non-Res.	47,100	41,500	152,700	2,911,000	877,900
Supermarket	Non-Res.	28,900	5,100	9,600	436,100	189,100
Warehouse	Non-Res.	142,400	12,100	37,900	389,000	91,800

[a] Non-Res. = nonresidential

[b] kWh = kilowatt-hours

24

Appendix B. Distributions in Cumulative DPV Capacity Estimates for 2050

To investigate the differences in results across different agent resolution levels more closely, we can look at the distributions of outcomes in 2050. Figure A depicts kernel density estimates of cumulative PV capacity for each agent resolution level. The figure suggests that at low resolution levels ($n \in \{1,2\}$), the distributions are heavily right-skewed and possibly bimodal.[11] As agent resolution increases, the distributions become less skewed. At the highest resolution level ($n = 20$), the distribution appears almost normal, with just a small degree of skewness.
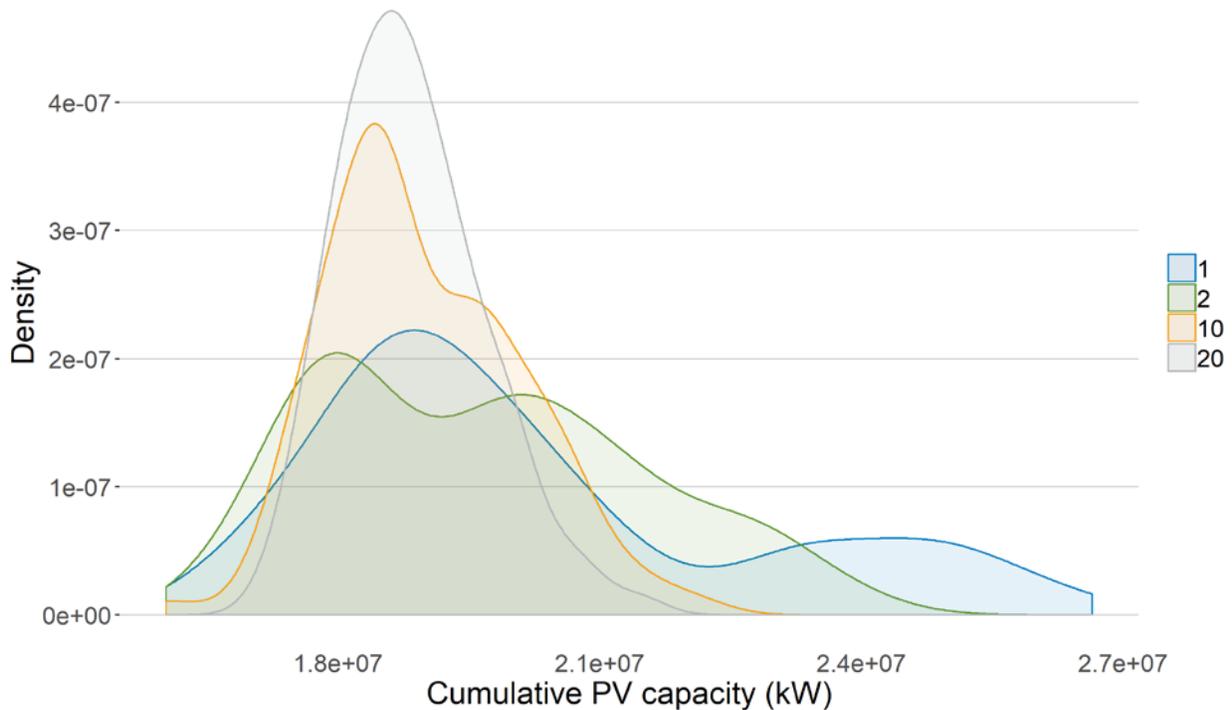
Figure B-1. Distributions of state-wide cumulative PV capacity in 2050 by agent resolution

---

[11] We say "possibly" bimodal because this feature depends on the smoothing parameters used to calculate the densities.

Appendix C. Annual DPV Estimates by Agent Resolution

We plot cumulative DPV capacity by year and agent resolution for the residential and commercial sector, as well as for both combined. Visually, the nonresidential sector more closely resembles the combined state-level results.
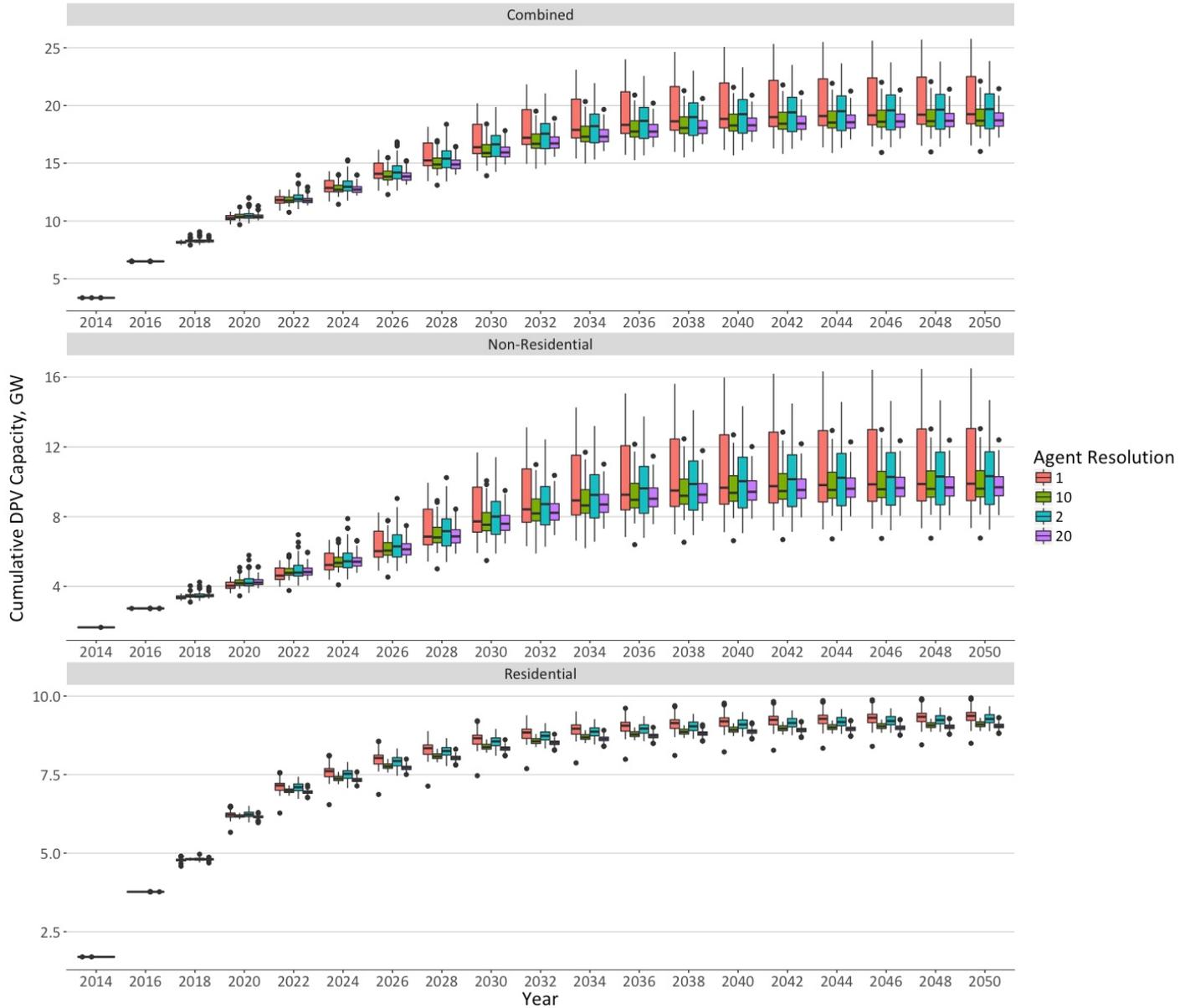

Figure C-1. "Boxplots" of cumulative PV capacity by year and agent resolution

Appendix D. Annual DPV Adoption by Agent Resolution

Here we illustrate the year to year trends in new DPV adoption by sector. Lines reflect the mean across 100 simulations at each agent resolution and ribbons reflect two times the standard deviation. We see mass adoption in the nonresidential sector delayed relative to the residential sector.
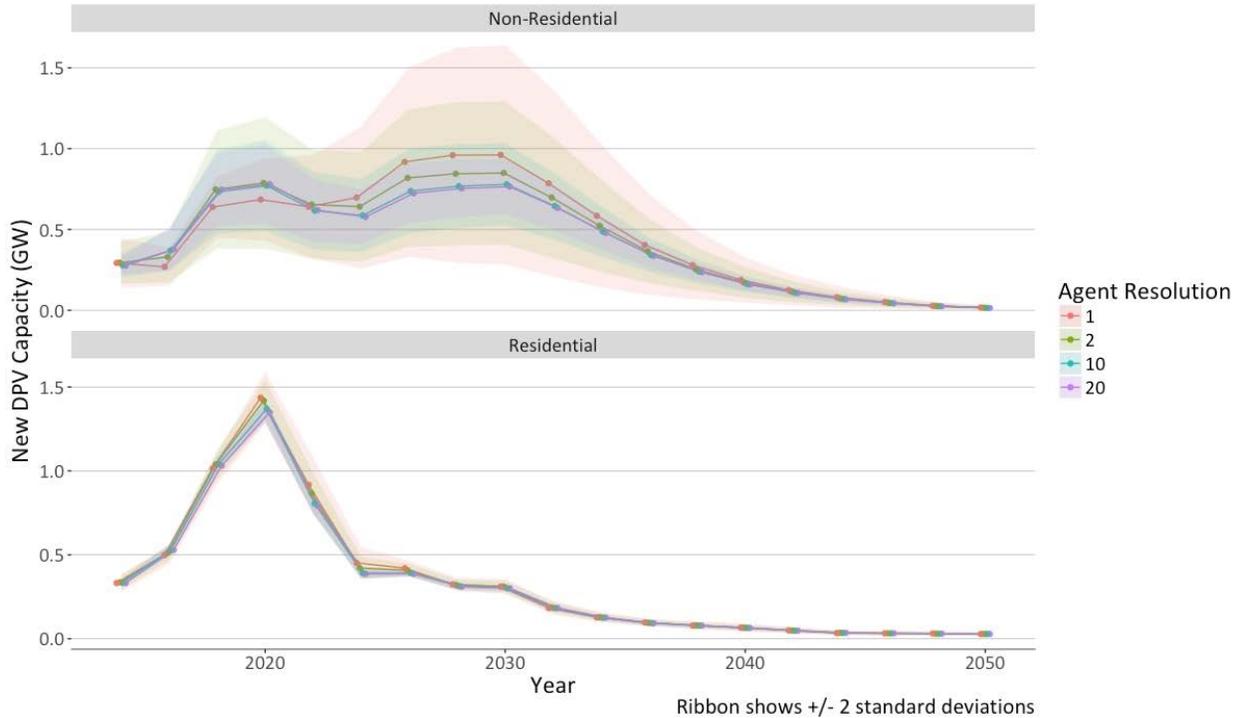

Figure D-1. New DPV capacity by year, agent resolution and sector

27

Appendix E. Example Illustrating Bias of the DPV System Sizing Mechanism

Here we illustrate how selecting the minimum of possible system sizes reduces average selected system size as agent resolution increases.

Consider first the simple case of a 1-agent resolution. In this case the selected size will simply be the lesser of the system sized to either load or roof area, or 10 kW resulting from the load.

Table E-1. Average System Size at 1-agent Resolution

Agent	kW sized to Load	kW sized to Roof Area	Selected Size, kW
1	10	15	10
Average Size	**10**	**15**	**10**

Next, consider the case of a 2-agent resolution and note that average sizing by load and roof area remain consistent. In this case, if the roof sizing is consistently larger than the load sizing, so the average selected system size remains consistent at its maximum of 10 kW.

Table E-2. Average System Size at 2-agent Resolution with Proportional Roof Sizes and Loads

Agent	kW sized to Load	kW sized to Roof Area	Selected Size, kW
1	10	15	10
2	10	15	10
Average Size	**10**	**15**	**10**

If roof sizing is not consistently larger than load sizing however, the average selected size must decrease as shown in Table #, where it decreases to 9.5 kW.

Table E-3. Average System Size at 2-agent Resolution without Proportional Roof Sizes and Loads

Agent	kW sized to Load	kW sized to Roof Area	Selected Size, kW
1	10	21	10
2	10	9	9
Average Size	**10**	**15**	9.5

At a 1-agent resolution, we are guaranteed to observe the maximum possible system size. If relative sizing is purely random, however, at a 2-agent resolution we would expect one sizing column to be consistently greater than the other only 50% of the time. This likelihood is reduced to 25% at a 3-agent resolution, and it continues exponentially decrease as the number of agents increases. Thus, intuitively we expect to observe an average selected system size to be below its maximum more often as agent resolution increases.