Energy 112 (2016) 342–352

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Data granularity and the optimal planning of distributed generation

L. Kools*, F. Phillipson

TNO, P.O. Box 96800, 2509 JE The Hague, The Netherlands

ARTICLE INFO

Article history:
Received 25 February 2016
Received in revised form
30 May 2016
Accepted 18 June 2016
Available online 5 August 2016

Keywords:
Stochastic programming
Distributed generation
Capacity planning

ABSTRACT

Research regarding the optimal planning of distributed generation is often based on coarse energy use and generation data, which does not accurately reflect real variations in energy profiles. This paper investigates the impact of this lack of temporal variation on the optimal planning of distributed generation. The problem of loss minimization in the residential setting is used as a guideline. The outcomes of a stochastic optimization model for energy profiles defined on different time aggregation levels are compared. At first glance, modeling on a finer time scale seems to affect optimal planning solutions, with a shift from variable stochastic sources to sources that provide constant generation. However, it turns out that the gains of using these new optimal solutions in terms of reducing energy losses are limited. The results suggest that for optimization purposes it is not necessary to use data at a resolution smaller than hourly time steps. If energy profiles are defined on time steps smaller than one hour it is important that the full range of the stochastic fluctuations is taken into account, rather than evaluating a couple of scenarios.

1. Introduction

In the recent years there has been an increased focus on 'smart' electricity systems that continuously monitor, match, and control energy use and generation in order to facilitate the implementation of onsite-generation [1–4]. In the context of these developments opportunities arise for intensive data collection. Whereas data used to be collected on the level of daily or hourly means, now data is collected and even used in real time [5,6]. Since household load and renewable generation are known to fluctuate on the level of minutes or even seconds [7], fine grained data collection could be relevant to provide an accurate picture of energy use and generation. Indeed these new possibilities are embraced by researchers who are formulating increasingly precise energy profiles [8–11]. However, when it comes to simulating or optimizing the grid, the use of accurate data comes at the price of computational efficiency. Adding more accuracy in the time dimension comes with the need to acknowledge the unpredictable nature of the short term fluctuations, which requires highly complex models. Moreover, intensive data collection and storage is not without costs. The question arises whether the possible improvements to distributed generation (DG) planning owing to more accurate data outweigh the costs of acquiring, storing, and using the data.

In this paper the impact of using electricity use and generation data of higher temporal resolution on optimal DG capacity planning is analyzed. To guide the analysis, a general model for determining the capacities of DG's at which energy losses are minimized is employed; an elaboration on [12]. The focus of this paper is not on the calculated optimal capacities itself, but on the influence of data choice on the calculated optimal capacities. First, the implications of high temporal data for modeling flexibility are outlined. Then the impact of increased data granularity on estimates for the performance measure and optimal capacity levels is evaluated by comparing the outcomes of the optimization model for energy profiles with different levels of time aggregation. The latter has drawn especially little scientific attention so far.

1.1. Granularity and system performance

Quite a few papers have evaluated the bias from the use of coarse data when calculating performance measures for the power distribution system. The resulting picture is ambiguous. As Bucher et al. [13] explain, much depends on which part of the system is being analyzed. For a realistic representation of maximum power, maximum voltage or energy flows, a one minute time step seems appropriate; whereas for evaluating transient currents and

* Corresponding author. Present address: Leiden University, Steenschuur 25, 2311 ES Leiden, The Netherlands.
E-mail addresses: l.kools@law.leidenuniv.nl (L. Kools), frank.phillipson@tno.nl (F. Phillipson).

http://dx.doi.org/10.1016/j.energy.2016.06.089
0360-5442/

voltages, smaller time steps are necessary. For broad overviews of energy flows more coarse data is sufficient. Moreover, whereas one hour time steps may be too coarse for analyzing data on the individual level (i.e. domestic loads), when looking at multiple households, much of the short term fluctuations are balanced out due to the aggregation of different profiles [14].

At least two papers focus specifically on the impact of granularity on measuring the mismatch of demand and generation (which is the input to loss minimization models). Wright et al. [15] use energy use information collected at a one minute time interval from seven different households, to calculate the degree of measurement error in export and import proportions when this data is averaged over five, fifteen, or thirty minute time intervals. They conclude that five minute time steps provide a good balance between accuracy and the burden of data size. Cao et al. [16] do not only consider the short run variability of demand, but also of on-site photovoltaic (PV) generation. They simulate energy use and generation profiles on a one minute interval and from there construct profiles on five, ten, fifteen, thirty, and sixty minute time steps. Using this input they compare on-site energy fractions and on-site energy matching for different setups of the site. They show that using one hour averages leads to large biases compared to the one minute data. However, they cannot give specific guidelines with respect to the granularity. The size of the bias depends on the setting evaluated, so that in some cases five minute data may suffice, while in others much can be gained from using finer resolution data.

1.2. Granularity and decision making

Even if one can assume that increased data accuracy leads to more reliable performance measurement of the power distribution system, this does not necessarily imply it improves capacity planning. The consequences of adding precision to the input of optimization models for DG capacity planning is rarely discussed in scientific literature. Ochoa and Harrison [17] provide a first step in the discussion by advocating the use of multi-period (‘energy’) models for loss minimization instead of the popular one period (‘power’) models. Power models evaluate the system at one moment in time, using a snapshot of the systems performance for optimization. Energy models, on the other hand, allow to evaluate the system over a span of time, thereby accounting for time variations and time dependencies in demand and supply. Ochoa and Harrison show that using one period models leads to a downward bias in the performance measure and a resulting overestimation of optimal DG capacity levels.

The next step is to determine the appropriate length for these time periods. There are two recent papers in which the impact of using time periods shorter than one hour in optimal DG capacity planning is discussed, both in the context of cost minimization. Hoevenaars and Crawford [18] investigate the optimal capacities of the elements of a standalone hybrid power system for different temporal resolutions of the data, ranging from one second time steps to one hour time steps. They explain that the appropriate level of analysis depends on the components included in the model. Whereas the optimal amount of diesel fueled generation is highly dependent on the time step chosen, systems including solar modules, wind turbines, or batteries are less dependent on the granularity of the data used. Hawkes and Leach [19] look at optimal sizing of a combined heat and power (CHP) unit for a single household and finds that the calculated optimal capacity based on the hourly data is twice the optimal size found when using data on a five minute level. The resulting difference in costs can be up to 16%. They believe it is not necessary to use data on an even smaller granularity, as the improvements from the level of ten minutes to five minutes is already rather limited.

1.3. A stochastic approach

When one wants to use high resolution data for decision making, the accompanying optimization model naturally becomes more complex. Not only does increasing the temporal granularity of energy profiles result in adding data points, also one now needs to acknowledge the unpredictable nature of short term fluctuations in energy use and generation, such that stochastic formulations are required. By adding granularity to the input of an optimization model one thus needs to switch from evaluating the performance of a system under a typical situation to evaluating the expected performance of a system given the complete range of possible situations. In this way the amount of information evaluated in the model increases exponentially and one must apply approximations to keep the model tractable. Not surprisingly, most models for optimal DG planning are defined in a coarse deterministic framework, as one can conclude from the extensive overview of models in Kumawat et al. [20]. They review and classify more than 70 high quality research papers on optimal DG planning that have been published after 2010. They distinguish the following four approaches to modeling the load as input for the model. Within the brackets the percentages of papers reviewed that fall within each classification are given.

1. one-load level with single case (16%)
2. multi-load levels (51%)
3. time-varying (practical system loads) (28%)
4. probabilistic generation considering uncertainties in load (5%)

Most of the examples mentioned, such as [21–24] use hourly, daily or even monthly time steps. Only three of the papers [25–27] consider probabilistic generation. All of these consider load per hour, just as the more recently published paper within this category by Kayal et al. [28].

In those applications dealing with the reliability of the power distribution system, elegant optimization models have been defined where higher time granularity does not add complexity, borrowing from techniques used in telecom research, see Refs. [29–31]. The focus then lies on the performance of the system under extreme conditions. Unfortunately, such an approach does not readily apply to the type of optimization problems where average performance of the system (as is the case for loss minimization) is the main concern.

The aforementioned research on granularity and decision making acknowledges the stochastic nature of the short term fluctuations by considering several likely data profiles (samples or snapshots) rather than one average profile. In this way the computational burden of the analysis can be kept to a reasonable level, without needing to compromise on the complexity of the model. The stochastic nature of short term fluctuations could impact optimal decisions, which may not be adequately captured using a small amount of samples. This paper aims to take the full span of possible realizations into account by framing the problem in the language of stochastic optimization. This allows for consideration of the complete range of stochastic fluctuations in the model at the cost of the level of detail that can be included in the model.

1.4. Outline of the paper

In Section 2 the problem used for the calculation is presented and the consequence of adding precision is explained with respect to complexity and computational effort. Then in Section 3 the parameters and data underlying the numerical calculations are