Supplementary Materials for

Data-driven modeling of solar-powered urban microgrids

Arda Halu, Antonio Scala, Abdulaziz Khiyami, Marta C. González

The PDF file includes:

Supplementary Information

S1 Cambridge and Pecan Street data set selection

To model the hourly demands of Cambridge after the Pecan Street users, we first look at the monthly total usage distributions of the two data sets. The monthly usage distributions of Cam- bridge users in the 200-2000 kWh range is very similar across different months. In Figure S1, we show the distributions for January and July as an example. The monthly usage distributions
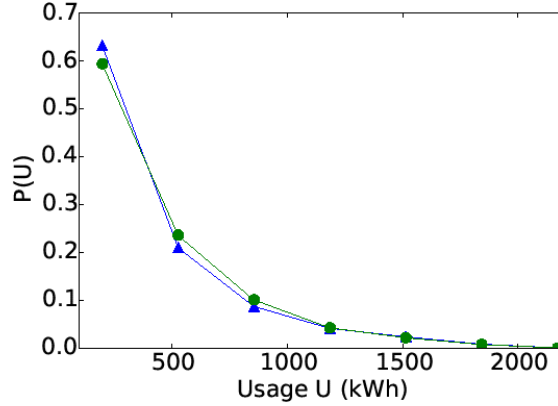

Figure S1: The monthly usage distributions of Cambridge for January (triangles) and July (circles).

of Pecan Street users, on the other hand, are different for summer and non-summer months, mainly due to the substantial increase in the cooling load during summer months. In Figure S2, this increase can easily be seen during during the afternoon for the summer months in the Pecan Street data set. We also note that the solar generation in the Pecan Street users does not vary considerably throughout the year (Figure S3), as the installation sizes and irradiation patterns are similar across residential users in similar areas.

Using the fact that Cambridge monthly usage does not change much for different months, and also to offset the seasonal difference from the much warmer Austin, we choose the July data set of Cambridge. We then systematically compare the distributions of Cambridge usage in July with the Pecan Street users throughout the year (Figure S4). We see the clear agreement in the non-summer data sets whereas the summer distributions are noticeably different. Hence, we base our modeling of Cambridge hourly demands on Pecan Street users by exploiting the fact that the distributions point to similar usage behavior between the two user base for non-summer months. Thus, in our simulations for the paper, we choose the July data set for Cambridge and the April data set for Pecan Street (5th panel in Figure S4).
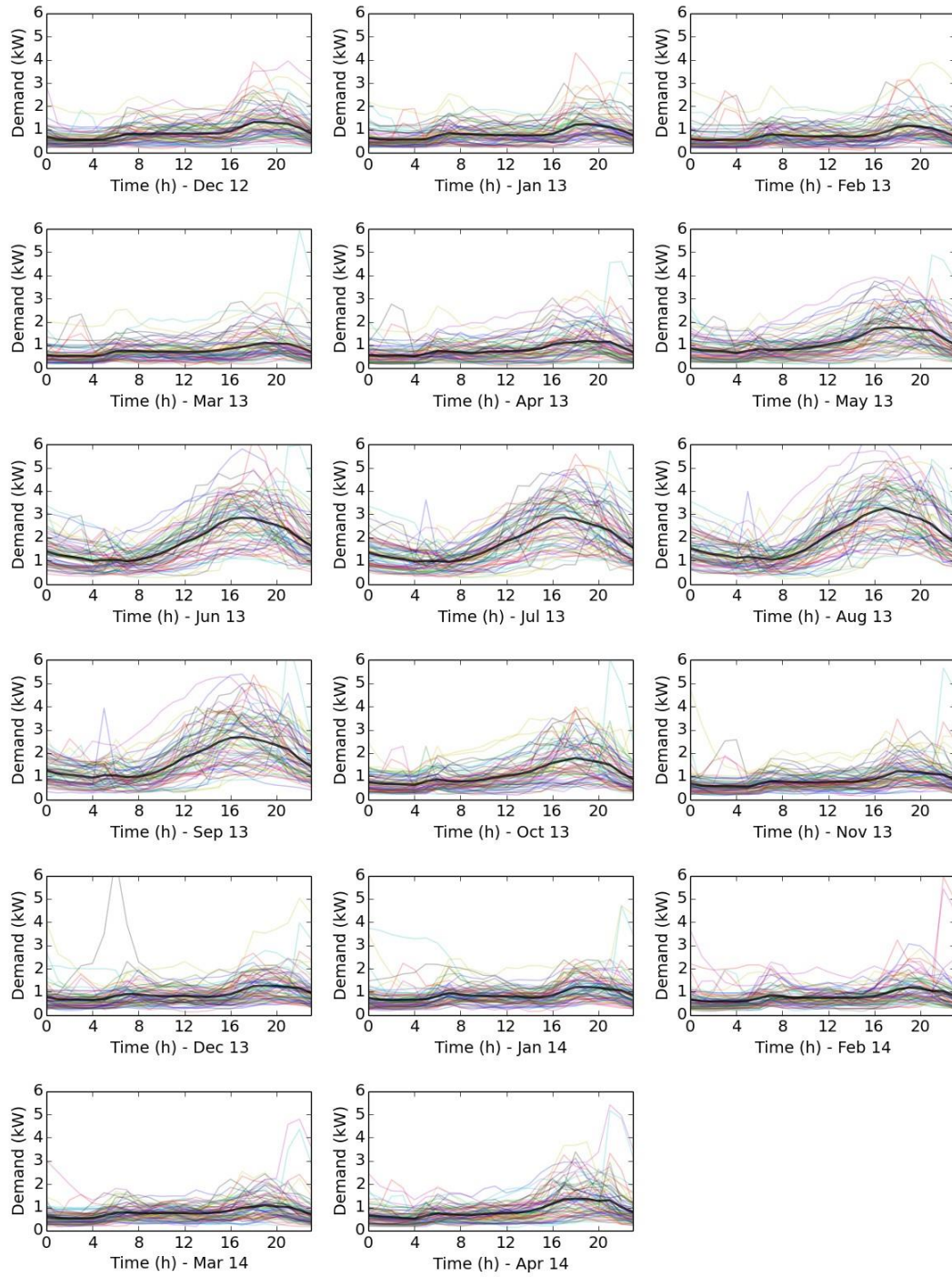
Figure S2: The hourly demand profiles of Pecan Street users for 17 months from December 2012 to April 2014. Solid line indicates the average over all users.
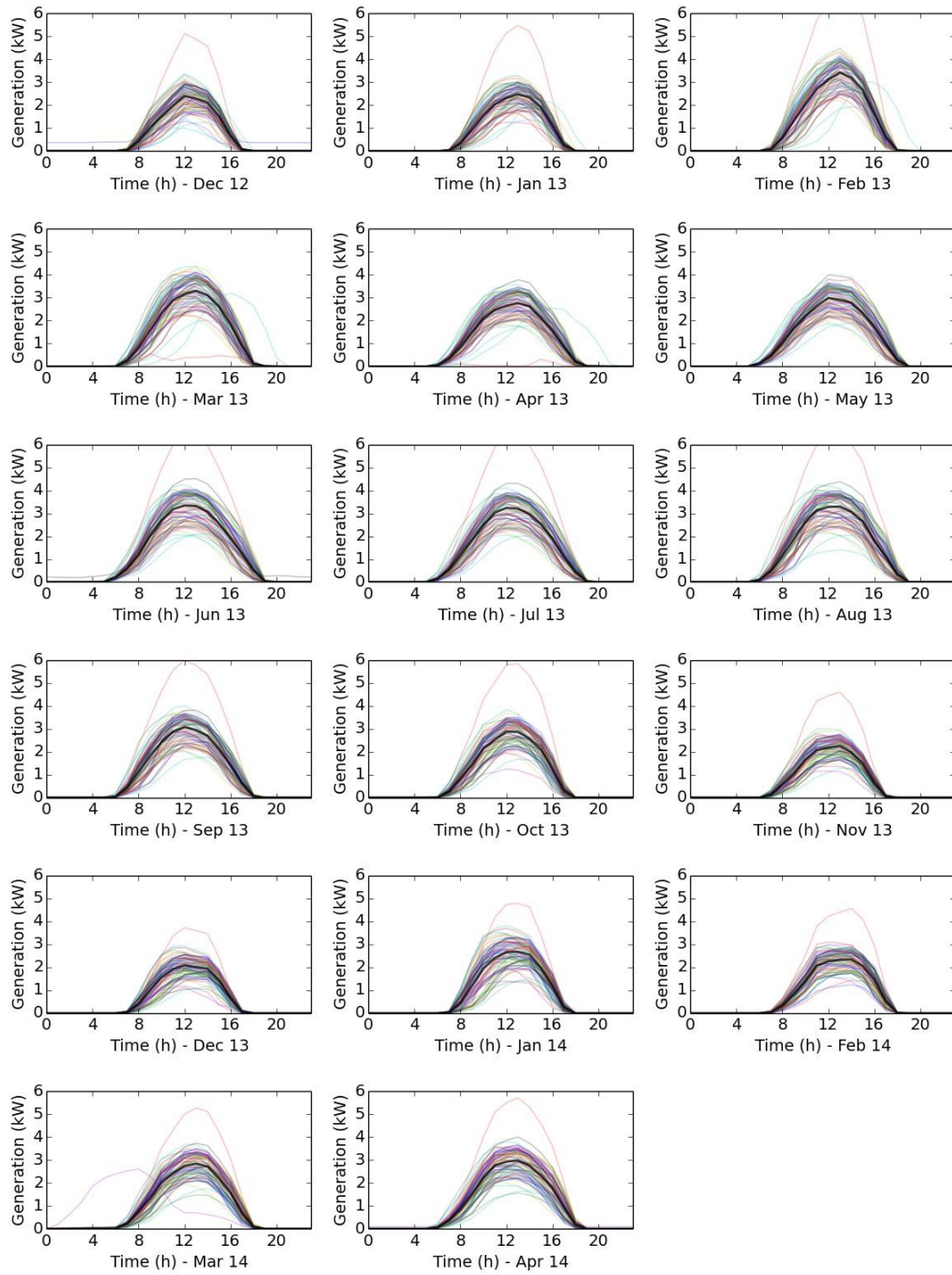
Figure S3: The hourly solar PV generation profiles of Pecan Street users for 17 months from December 2012 to April 2014. Solid line indicates the average over all users.
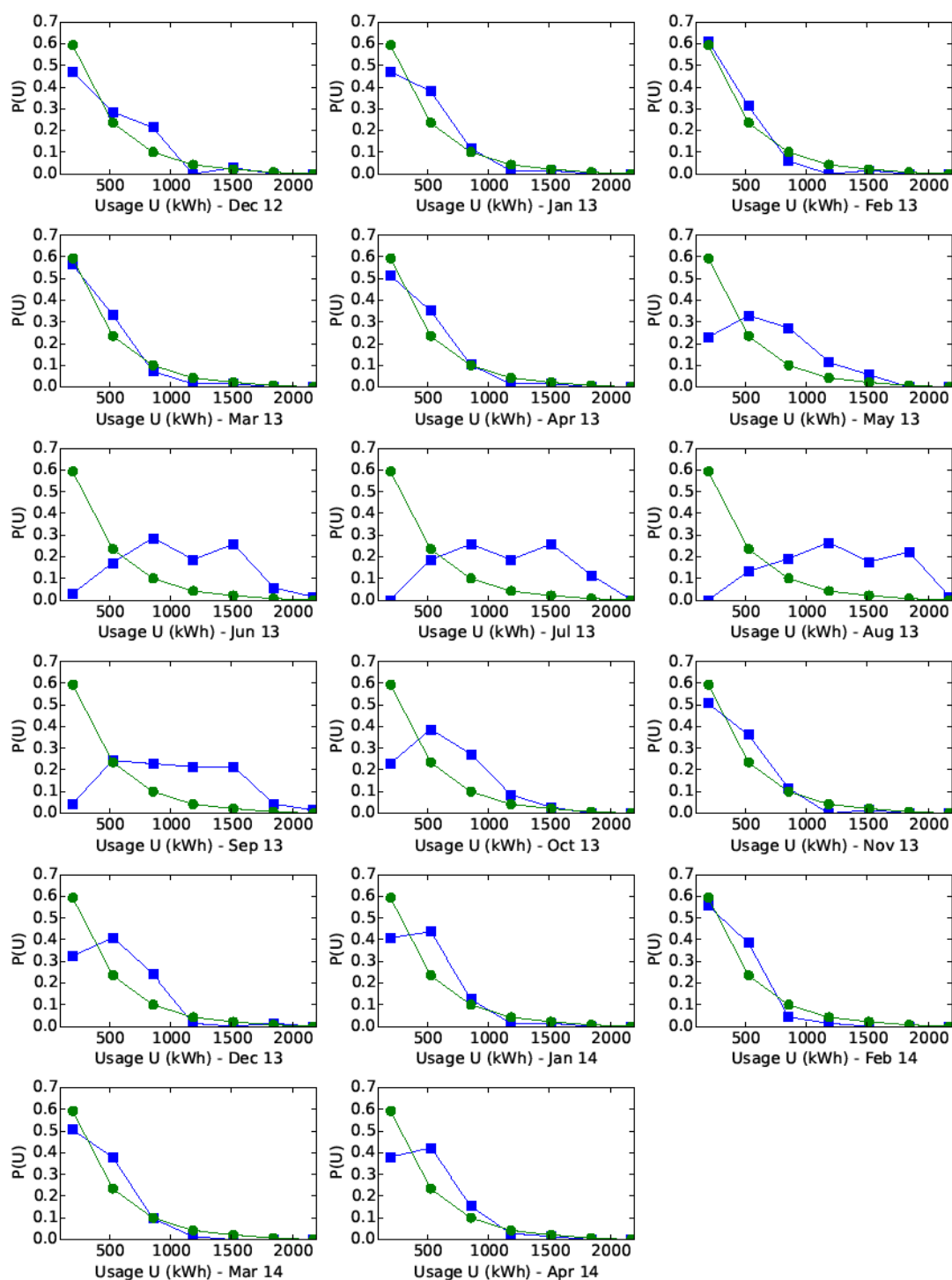
Figure S4: The monthly usage distributions of Cambridge in July (circles) and Pecan Street for 17 months (squares) from December 2012 to April 2014.

S2   DC versus AC power flow

We use DC power flow throughout our analyses as it is a simple, linearized and fast way to compute active power flows. However, we find it important to remark on the basic assumptions of the DC approximation, in particular the high reactance to resistance ratio (X/R) requirement. For this purpose, we study the behavior of the active power flow P as a function of X/R under DC and AC power flow calculations. Following [43], we quantify the discrepancy between the active power flows calculated by the AC and DC power flow methods in terms of the percentage error P_{err}, defined as

$$P_{err} = \frac{P_{AC} - P_{DC}}{P_{AC}} \times 100.$$

Furthermore, we make this sensitivity analysis specific to the microgrid case. We use a basic radial topology inspired from current real-world and laboratory-scale microgrid experiments tailored for applications such as apartment complexes and high rise commercial buildings where there are many users per building [44]. This topology consists of a backbone DC line running between the PV module and the PCC, with branches serving each unit on each floor (Figure S5).
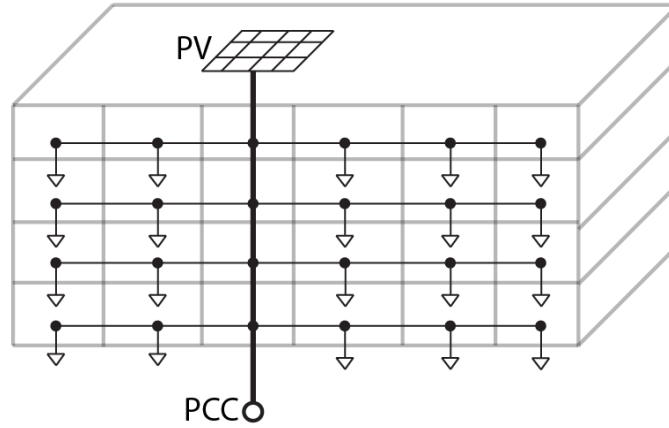

Figure S5: The topology of the building microgrid used in the power flow sensitivity analysis.

    In line with our main approach in the paper, we use real hourly demands from Cambridge, which comes from a part of our data set that has multiple users per parcel. We choose a parcel that has 20 users for our sensitivity analysis and use their hourly demands at 1pm. With this setup, we calculate the active power flows using DC and AC power flow approaches. For the various branches in the system having $R = 5m\Omega$, $R = 2.5m\Omega$ and $R = 0.17m\Omega$, we vary the resistance of all branches from 0.1*R to 10*R for X values between 0.1 to 125 and calculate the DC and AC power flows for the resulting X/R ratio. In Figure S6, we show the results for the small X/R regime between 0.5 and 12. We see that the percentage error has the same asymptotic behavior as demonstrated in [43] and is never more than a few percent, even when the reactance to resistance ratio is as low as $X/R = 0.5$. This is in agreement with recent studies showing that DC power flow is on average wrong by a few percent with respect to the computationally intensive AC power flow [45, 46]. Moreover, the probability distribution of flows in the networks is statistically similar for AC and DC calculations [46].
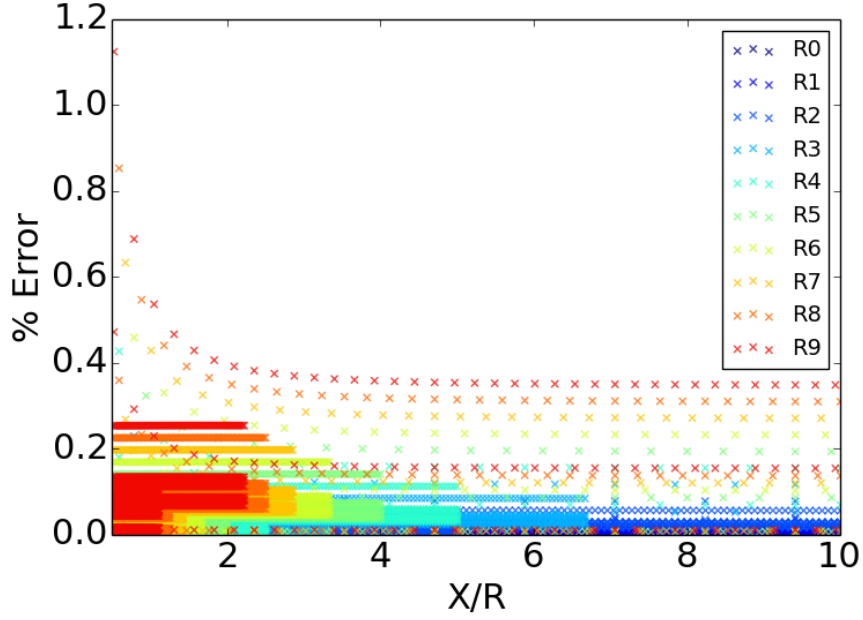
Figure S6: The % error (P_{err}) between AC and DC power flow as a function of the *X/R* ratio for the proposed topology for different values of *R*.

S3   Microgrid size distributions

We measure the size of microgrid *M* as the number of users N_M in the microgrid. Our k-means partitioning into microgrids, described in the main text, yields microgrids of relatively uniform size. To quantify the variance in the microgrid sizes, we plot the microgrid size distribution in Figure S7. We see that it follows a normal distribution with μ = 23.6 and σ = 9.2 for 200 microgrids in Cambridge.

S4   Theoretical q_c as a function of α

To calculate the theoretical value of q_c, we use the formulation devised in [47] for randomly connected graphs with no loops and a given degree distribution $P(k)$, which states that $(k^2)/(k)$ = 2 at the critical point . We calculate q_c^{th} for our optimized networks over the entire range of α and observe that it falls within the range 0.20-0.40 (Figure S8).
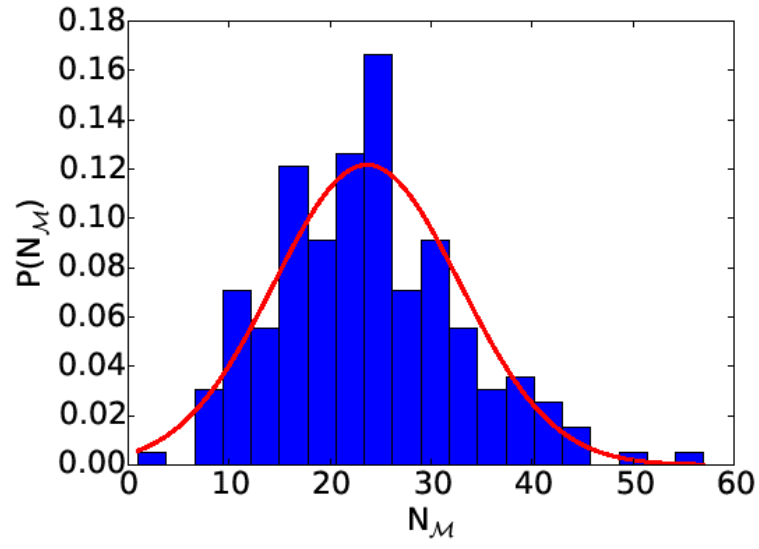
Figure S7: Size distributions of the microgrids in Cambridge partitioned using *k*-means.
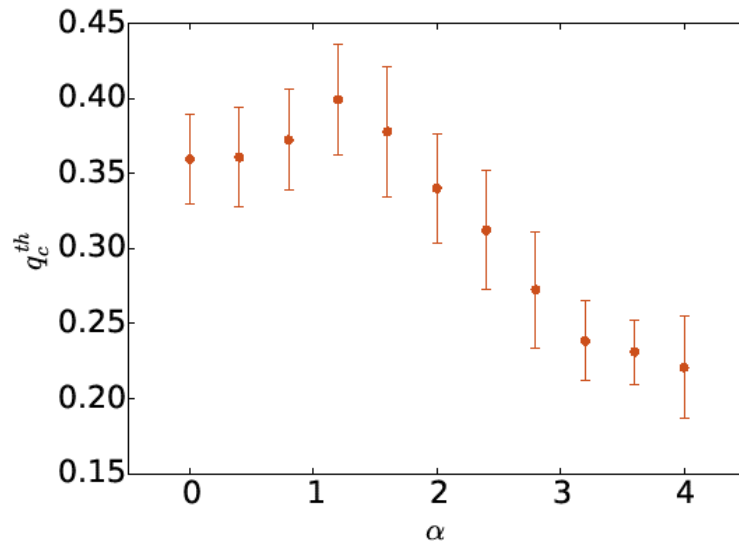

Figure S8: Theoretical q_c calculated over the whole range of α values, averaged over 10 realizations.