Appendix

A Algorithm for the percolation on the intersection points

The algorithm is based on the geographical location of intersections. We consider a pair of intersections as connected if they are no more than d meters apart. In order to reduce the computational complexity of the procedure, the actual analysis is performed using a grid of squared cells (10×10 meters each). A cell has one of two values: 1 if at least one intersection is within its area or *null* if it contains no intersections. As the percolation analysis is based on distance, we calculate a distance grid where each cell is assigned the distance to the closest cell that contains an intersection. We use this grid in the percolation procedure.

The percolation procedure for a distance d consists of the following steps:

1. Each cell of the distance grid that has a distance value of d meters or below is marked as 1, otherwise, it is marked as *null*.
2. A unique identifier is assigned to each continuous set of marked cells. A cell is considered adjacent to its four nearest neighbours (i.e., its von Neumann neighbour).
3. Each intersection is assigned the unique identifier of its containing cell.

The method is implemented in ESRI ArcMap 10.1 using the following tools:

- The intersection grid is created using the *Points to Raster* tool.
- The distance grid is created using the *Euclidian distance* tool.
- The marked cells grid is created using the *Raster Calculator* tool.
- The unique identifiers grid is created using the *Region Group* tool.
- The unique identifiers are copied to the intersection points using the *Extract Values to Points* tool.

B Algorithm for the network based percolation

Given a graph of the road network, where nodes represent intersections and the weight for each edge is the length of the street that connects them and a certain metric threshold (e.g. 5000m) we produce a network percolation via the following steps:

1. We select the link of the graph with the smallest weight (distance), generating a new cluster and inserting both its nodes into the cluster.
2. We keep a first-in first-out queue of *nodes to expand*, from which we extract a node to continue the process. We add both nodes of the link selected in step 1 to this queue. Nodes are only added to this queue if they are not already included.

3. We extract a node from the queue of *nodes to explore* and if a link departing from that node (not yet included in the cluster) is smaller than the threshold, include the link in the cluster and the end node of the link in the queue of nodes to explore.

4. We repeat step 3 until no further node can be expanded (the queue is empty) and if there are links left in the graph that do not belong to any cluster, generate a new cluster by choosing the smallest available link and repeat from step 1.

C Details of the clusters at the transitions

At each of the critical distances defining a transition, a set of clusters appears. There can be thousands of these, hence we opt to only visualise the 10 largest ones, setting colours representing a rank size: red is the biggest, blue is the second biggest, green the third etc. The rank and colour are illustrated in Fig. 4.

As discussed in the main text, although the critical distances differ for both systems (the percolation on the intersection points and the percolation on the network), see Fig. 1, the results are very similar. The maps in Fig. S1 and Fig. S2 represent the transitions for the percolation on the intersection points and on the network respectively.

D Correlation measure of the network percolation clusters and the urban area

For a given percolation result on the network, we categorise the type of each intersection as either being urban or non-urban. We define an urban intersection as an intersection that belongs to a cluster that is larger than Smin = 50 while the rest of the intersections are considered to be non-urban.

We use the polygons defined in the Corine dataset as a reference point. We generate a grid of 1km per 1km over the whole territory of the UK. For each square of the grid we assign two values, the first value is the area of the polygon that corresponds to the Corine cluster that intersects the square of the grid, the second value is the mass (the number of intersections) of the percolation cluster that has more intersections in the square of the grid. In order to be able to compare both systems we perform two types of analysis. The first is a Pearson Product Moment correlation between the values assigned to each square of the grid when there is both a cluster from the Corine and a percolation cluster. Fig. 8 shows that the highest correlation, for $R^2 > 0.7$ is given for the range of distances 300m to 400m. The second type is a measure of error comparing both. The procedure is as follows: when the squares of the grid do not have both types of clusters, we count the number of squares that have a Corine cluster but not a percolation cluster; and also the other way round, we count the number of squares that have a percolation cluster but not a Corine cluster. Finally we add both counts to get the total number of squares that do not have coincident clusters. The result is given in Fig. S3, and this shows that the total number of non-coincident clusters is minimised for $d = 300$m.

E Average cluster size and cluster distribution

A traditional approach to detect transitions in percolation processes, is to look at the average cluster size removing the largest component. In order to avoid very small clusters that hold no information with respect to the hierarchical structure of the urban system under consideration, we impose a minimum cluster size. We select this minimum size to be $S_{min} = 600$ intersections, since it gives enough resolution, and includes very small settlements. To put this number into context, the number of intersection points in large cities is of the order of 10^5 and of 10^4 for the 30 largest ones.

The results for both methods are given in Fig. S4. Given the multiplicity of transitions arising from the different merging processes, the curve presents many different peaks. The different sizes of the second largest clusters after transitions take place, obscure many of the transitions that take place in the system. An overall picture becomes easier to grasp if one looks at the evolution of the largest cluster size, as done in Fig. 2.

Let us now look at the distribution of cluster sizes. We investigate whether these are power law distributed. We use the method developed by [1], where a power law can be ruled out if $p \leq 0.1$. Note that $p > 0.1$ does not guarantee that the distribution follows a power law. We compute the distribution for clusters that have at least 1000 points, and we remove in all cases the largest cluster, so that the giant cluster is never considered. Given that we have a multiplicity of transitions, the second largest, and sometimes the top largest ones, can still be quite large compared to the rest of the clusters, especially for large distances. The results, up to distance $d = 760m$ are presented in Fig. S5. The cumulative distributions for some of the distances are given in Fig. S6. At the transitions, we note that a power law cannot be rejected. For small distances, we observe that around the transition, the cluster sizes are power law distributed. There is a clear region of distances after cities formed and the next transition occurs, at which the sizes are not distributed as a power law. The merging mechanism leading to the multiplicity of percolation transitions translates into a fluctuating exponent of the system. It is important to remember, that exponents arising from power laws are always very sensitive to the sample considered in the distribution, the number of events etc. In this case, we only consider clusters that have at least 1000 points and we always remove the largest one. The number of clusters hence vary enormously from distance to distance. In addition, one could argue that given the multiplicity of transitions, removing the largest cluster still leaves us with very large remaining ones that will be responsible for the next transitions; this is most evident for large ds.

In conclusion, the value of the exponent does not define the threshold at which cities are defined, nor whether the system is composed of cities or regions. It is important to recall that an urban system obeying a perfect Zipf law will have an exponent around 2 for the cumulative distribution of population size. The observed fluctuations hence contribute to the debate of whether cities are universally distributed according to a Zipf law or not, and certainly tell us that an urban system distributed according to a Zipf law does not necessarily represent cities. One might argue that cities distributed according to Zipf's law is a necessary but not a sufficient condition.

F Multifractal spectrum, dimensions: D_0, D_1 and D_2

In this section we provide a summary of the main mathematical expressions to compute the fractal dimensions D_0, D_1 and D_2. This is an extract from [2], where we give a detailed review of the methodology employed in order to compute these three measures from the whole multifractal spectrum. In summary, a (mono-) fractal has a measure that is homogeneous in space, such that within a small region ϵ

$$\mu(\epsilon) \sim \epsilon^{-D} \tag{S1}$$

where D is the fractal dimension. For a multifractal, μ is no longer homogeneous, and hence for each region i we can define a distribution function $P_i(\epsilon)$ of the measure

$$P_i(\epsilon) \sim \epsilon^{\alpha_i} \tag{S2}$$

where each subdivision of the space i has a value α_i. A fractal dimension $f(\alpha_i)$ can then be associated for the set of regions with the same value. The moments of the distribution function are obtained through the function

$$Z_q(\epsilon) = \sum_i P_i(\epsilon)^q \sim (\epsilon)^{-\tau(q)} \tag{S3}$$

The exponent $\tau(q)$ can be written in terms of the generalised fractal dimension D_q as

$$\tau(q) = q\alpha(q) - f[\alpha(q)] = (q-1)D_q \tag{S4}$$

where

$$D_q = \frac{1}{q-1} \lim_{\epsilon \to 0} \frac{\log_{10}[Z_q(\epsilon)]}{\log_{10}(\epsilon)} \tag{S5}$$

defines the whole spectrum for $q \in (-\infty, \infty)$. The 3 fractal dimensions are hence obtained for $q \in \{0, 1, 2\}$. For a monofractal, D_q is a constant for all qs. For $q = 0$, the fractal dimension D_0 corresponds to the dimension obtained through a box-counting algorithm. For $q = 1$, we get

$$D_1 = \lim_{\epsilon \to 0} \frac{-\sum_i P_i \log_{10} P_i}{-\log_{10}(\epsilon)} \tag{S6}$$

which has a very similar form to Shannon's entropy, and this is why this is called the information dimension. Finally, for $q = 2$, the dimension D_2 takes the form

$$D_2 = \lim_{\epsilon \to 0} \frac{\sum_i P_i^2}{\log_{10}(\epsilon)} \tag{S7}$$

and this in our case gives the correlations for pairs of intersection points to lie within the same box ϵ.

References

[1] Clauset A, Shalizi CR, Newman MEJ. 2009. Power-Law Distributions in Empirical Data. *SIAM Review.* 51(4):661–703.

[2] Murcio R, Masucci AP, Arcaute E, Batty M. 2015. Multifractal to monofractal evolution of the London's street network. *Phys Rev E.* 92(6):062130.

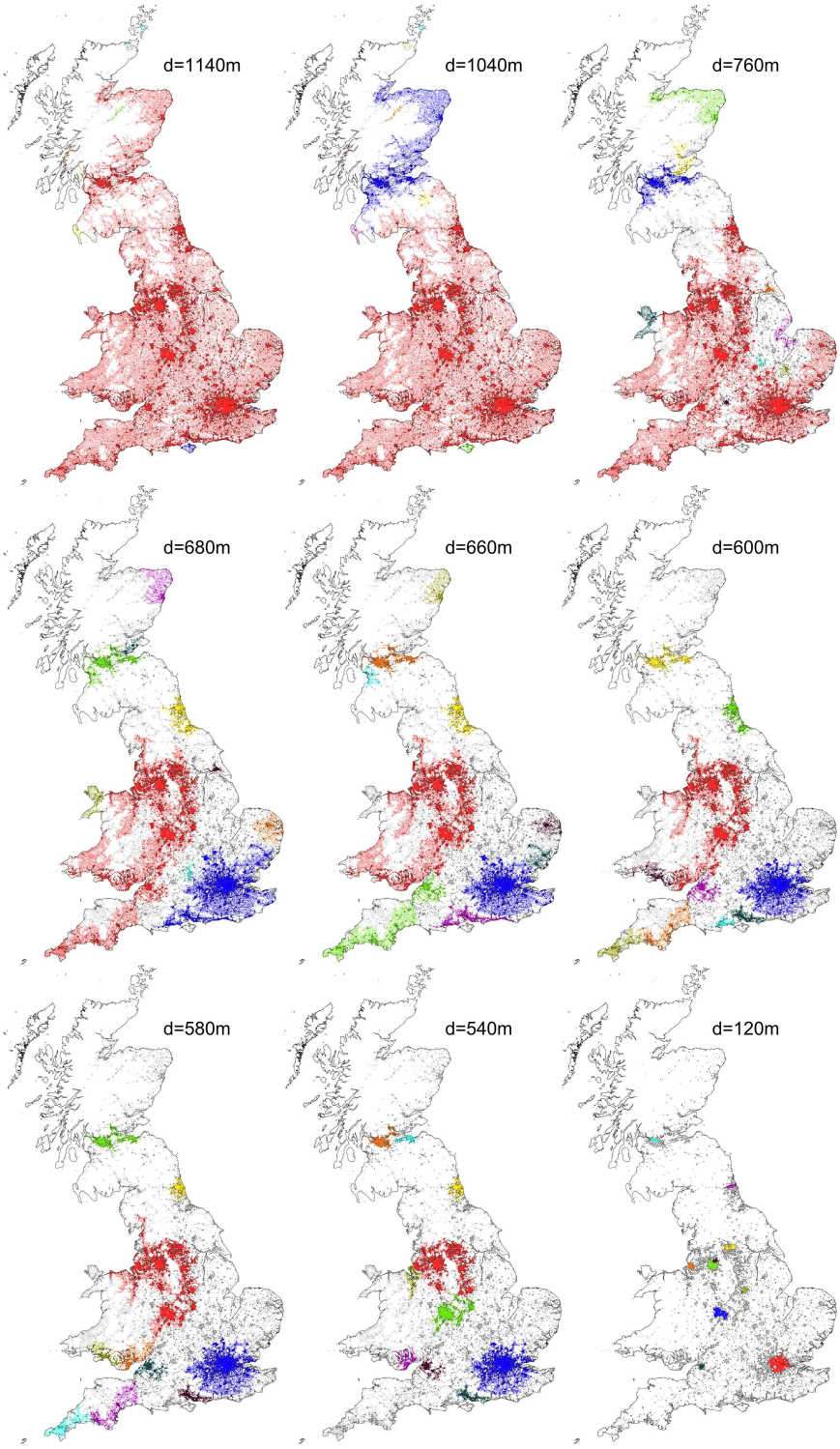

Figure S1: Maps of clusters at the transitions for the percolation on the intersection points. Only the 10 largest clusters have colours following the legend in the hierarchical tree.

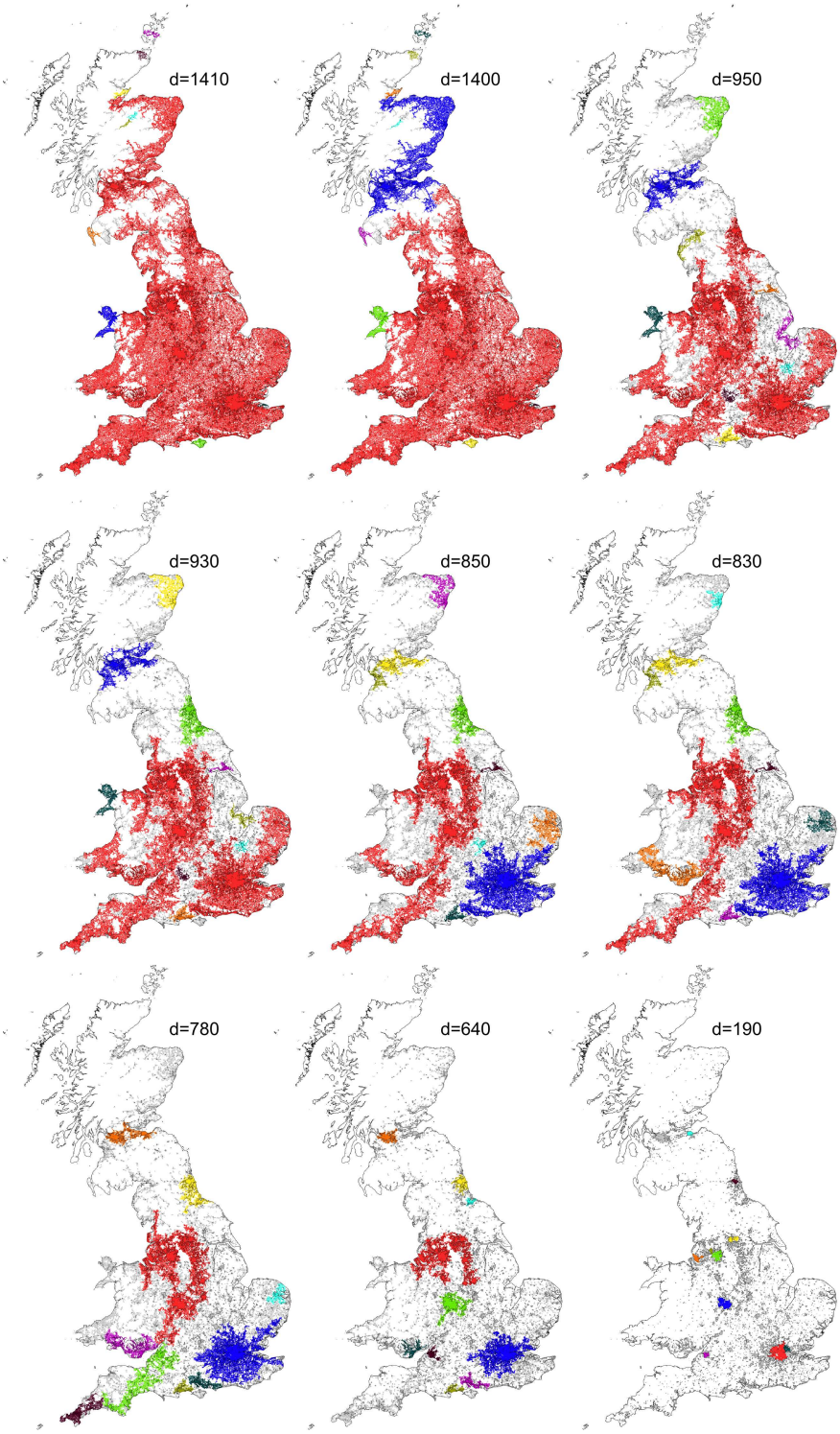

Figure S2: Maps of clusters at transitions for the network percolation. Only the 10 largest clusters have colours following the legend in the hierarchical tree.

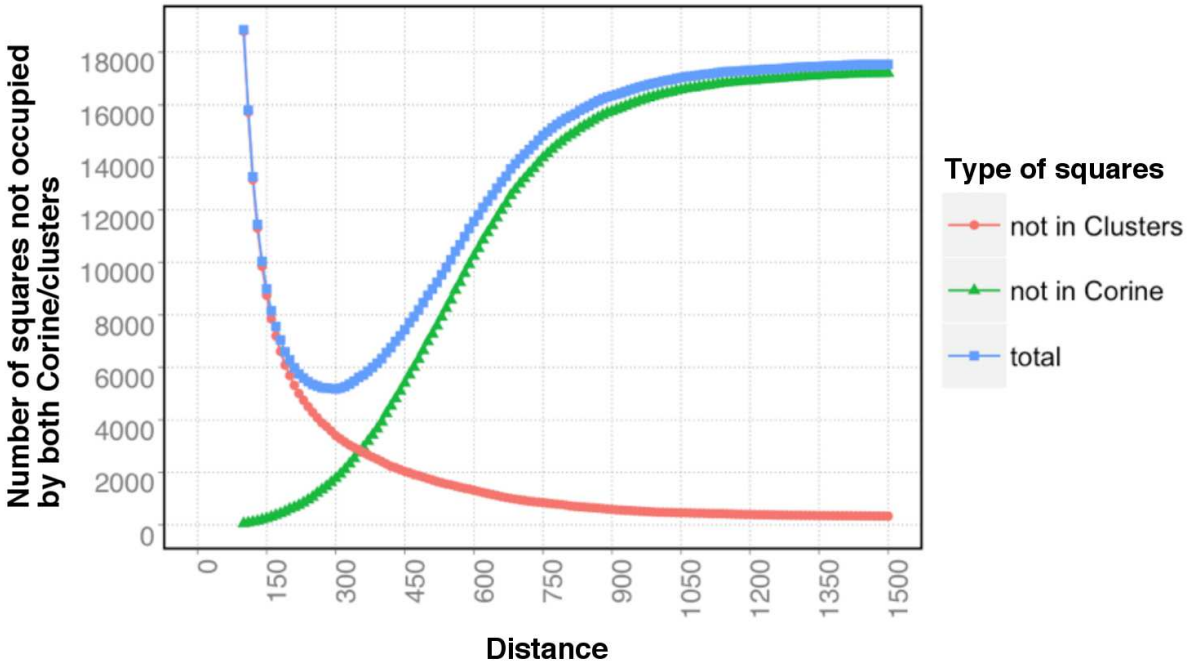

Figure S3: Measure of error of concurrency between the network percolation clusters and the urbanised clusters according to the Corine dataset.

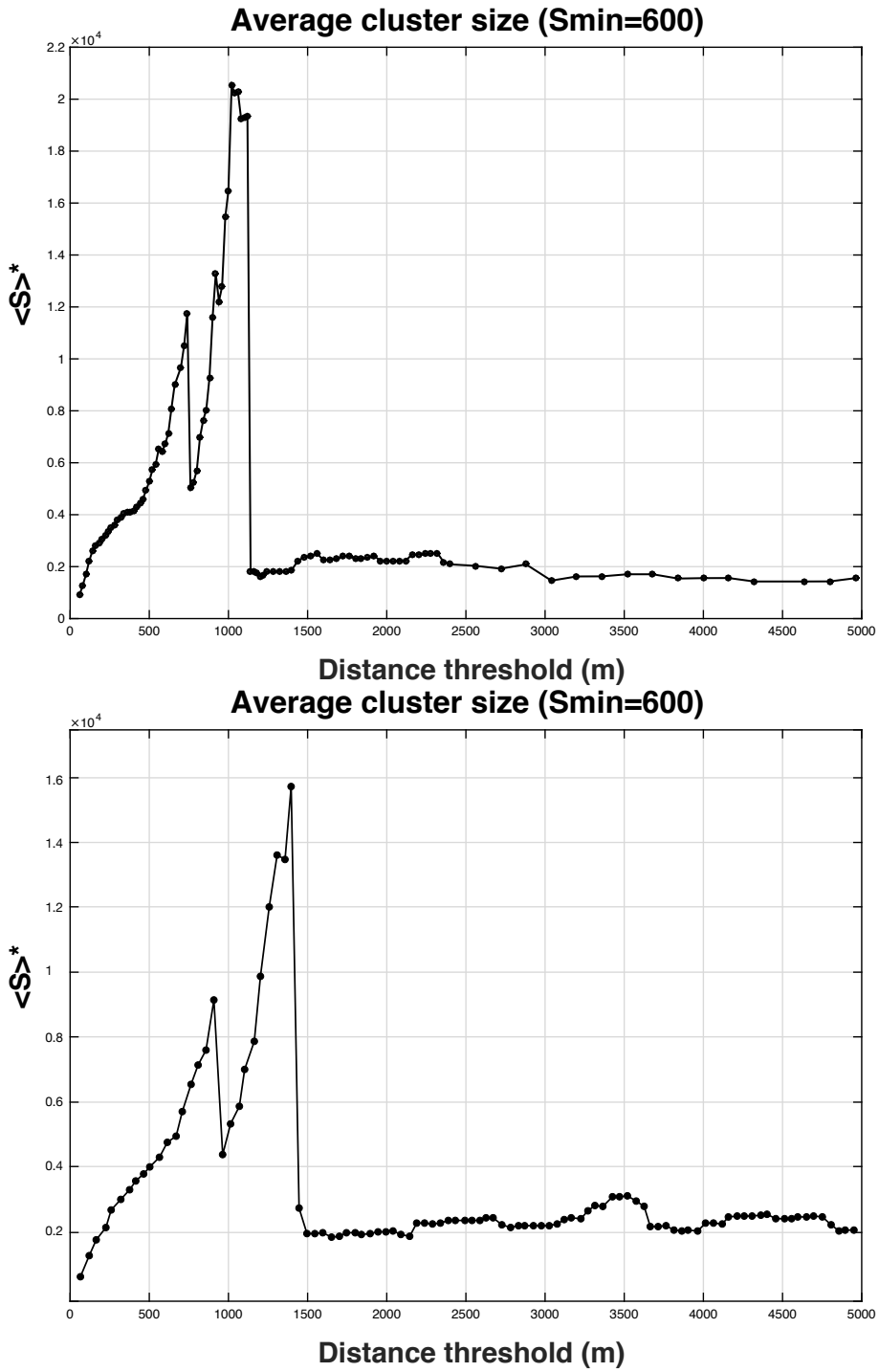

Figure S4: Evolution of the average cluster size removing the largest cluster, and including only clusters with at least 600 intersections. Top, plot for the percolation on the intersection points; bottom, plot for the percolation on the network.

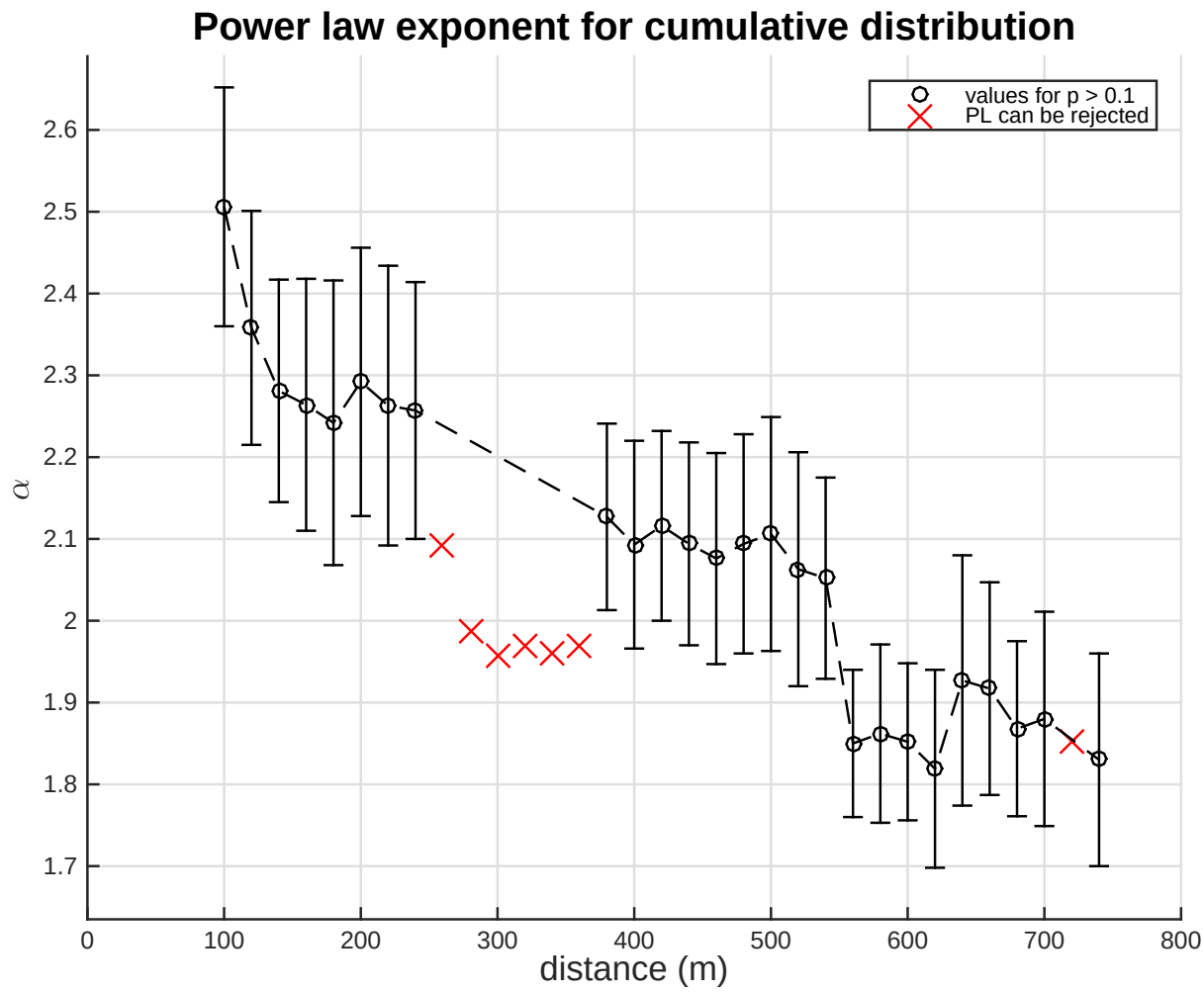

Figure S5: Values with confidence intervals for the power law exponent of the cumulative distribution using the method by Clauset *et al* in [1]. If $p \leq 0.1$ the power law can be rejected, and these are the red crosses.

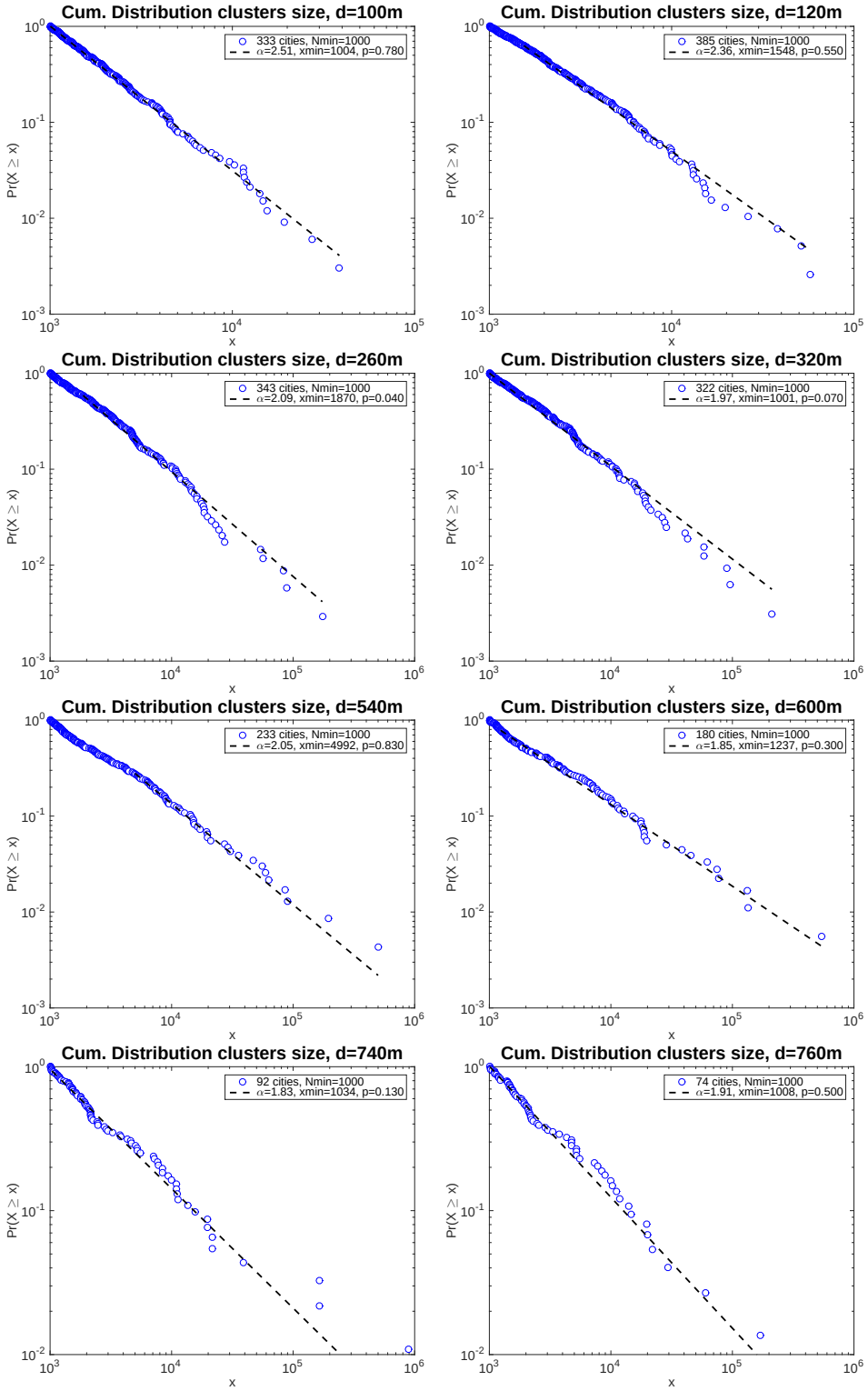

Figure S6: Cumulative distribution of the clusters size removing the largest cluster, and using the Clauset *et al* method [1].