

Urbanization, Energy Consumption and Entropy of Metropolises

Syed Amaar Ahmad

Senior Member, IEEE

syed.a.abmad@ieee.org

Metropolises are complex systems comprising social networks, engineering systems, agricultural output, economic activity and energy components. Urbanization stems from both increasing human population levels and expansion in a city's surface area. In this paper, we develop a model of how the population and area of a city affect its energy consumption patterns. We also show that with urbanization of smaller cities, there is a corresponding increase in the entropy (or variance) of the energy consumption, whereas, for larger cities, the entropy actually diminishes. This result interestingly implies that as metropolises scale, we may have a way to determine the point of a city's stagnation or decline. In the empirical analysis, we use power-demand data from the island state of Singapore between 2004 and 2019 to illustrate that with population growth, there is a tipping point for variance (or equivalently entropy) in the power demand. Moreover, we also provide a theoretical framework on how population growth and area growth cycles are mutually dependent and expand on how much information can be extracted about the entire city from a sub-region. Finally, we show that if the city's economic size (domestic product etc.) is proportional to the consumed energy, then for a constant population density, the economy scales linearly with the area. Our effort in developing a metropolis growth model is motivated by the need to understand how human behavior and organization at scale affect sustainability and economic growth.

Keywords: complex systems; Pareto-optimality; topological invariance; multi-scalar information; cellular automata; Dyson civilization

1. Background

An urban metropolis represents a large city or population concentration that is a significant economic, commercial, political and cultural center within a country or region and has vital regional or international links. A metropolis therefore represents a sprawling and unpredictable structure that is a complex system [1]. Capturing the features of such systems is often difficult, and appropriate mathematical models are needed to understand their workings. Consisting of social

networks, engineering structures, energy generators and urban agriculture, a metropolis is therefore an ultimate form of a complex system.

The growth of metropolises in terms of population, occupied area and economic activity is of critical interest to urban planners, governments and businesses. The population growth of cities is economically important in itself, helping to drive the national growth by virtue of sheer magnitude. Investments in construction, housing, infrastructure and energy sources help accommodate the rising population of cities. The economic and urban development literature on the connection between the growth of cities, population, spatial expanse (i.e., surface area) and per capita income has been crisply summarized in [2]. This work reviews key theories for urban growth that build atop the seminal work *The Isolated State* by nineteenth-century German economist von Thünen [3].

Urban development researchers have long known that transportation costs are a basic factor of the population of cities and underlying patterns of land usage [2]. A *linear monocentric* urban city is based on a model where commuting costs increase linearly with distance to the city center for workers. The literature for economic analysis on urban centers uses this monocentric model for housing and transportation. The results it draws also assume that the workers in the population are homogeneous and perfectly mobile. As [2] argues, the studies on the growth of cities so far have not been able to address how growth and technological progress mutually affect each other.

In this paper, we study a metropolis as a *complex system* [4] from a new perspective, where large-scale system behavior emerges from the underlying behavior of individual small-scale components. In complex systems, individual small components have behaviors that can aggregate to a more complex model. A metropolis exhibits a complexity where individual members of a large population move across its region under arguably some simple rules.

A common theme for complex systems is how they *scale*. The notion of scaling may have various connotations, ranging from wireless systems [5] to micro-organisms, and from gas thermodynamics to human civilizations [6]. Another aspect of complexity pertains to self-similarity, which is a recurrent theme in biological systems [7]. In this paper, the term *scale* is used to describe three situations: (1) the growth in the population size given a constant metropolis spatial region; (2) increase in the number of underlying spatial regions given a fixed population size; and (3) increase in the number of physically isolated (or otherwise) subsets of individual units given an overall constant population size and spatial region.

Existing work on the energy consumption of urban population centers uses empirical data for a sample set of US Zip codes (see [8] and its references). The study in [8] investigates the effect of socioeco-

conomic and demographic patterns on residential energy utilization patterns. It uses regression to show that total population and energy use are strongly correlated and that areas with low population densities tend to consume significantly more energy per capita than those with high population densities. The empirical relationship between the economic size and energy consumption has also been explored in [9, 10].

The energy consumption of the subregions of a metropolis can also help identify the internal dynamics and economy of the city. A key question is how representative the energy consumption of a subregion is for the remaining portion of the city. Self-similarity [4] sheds light on the fact that complex systems exhibit patterns where a small component may provide representation for the larger entity. Determining the energy consumption characteristics of a metropolis may also be useful in determining the economic dependencies or self-similarity of subregions of the city.

A complex system's internal function is exhibited by the information between its subcomponents, each operating with a set of simple rules. Developments in information theory [11] provide a powerful tool for quantifying the complexities and internal dependencies of systems. The *mutual information* reflects the dependencies between various subcomponents of a system. Each subcomponent provides inferences about the system's dynamics, based on statistical or probabilistic relationships.

The work in [5] explores the concepts of *multi-scalar information*, in terms of how a system's joint information is spread over each of its irreducible smaller components. It shows that there is a marginal utility of information depending on the scale. It is also concerned with whether the system's components (e.g., intelligent machines, particles or social groups, etc.) are largely independent or strongly interdependent.

1.1 Contributions

We treat the metropolis as a model with only two variables: population and energy-consuming *resource cells* (or simply cells), which could be a set of housing units, buildings, city blocks or arbitrarily large areas. A cell also represents some arbitrary geographical area, land resources or composite residential and industrial infrastructure. The consumed energy of each cell is presumed to be a hybrid mix of solar, fossil fuel or alternative sources. In this paper, we present the following new concepts:

1. We propose a topological model of the population distribution in a single metropolis. The population units are mobile and move back and forth between the cells. We treat the population distribution of the city as a random field; with an *Euler characteristic*-like function, we can represent the overall energy consumption of the city

fluctuating with the population distribution. The randomness stems from the underlying actual movement of people and goods throughout the city. As the population moves about the metropolis, the cells with the densest population change. In this model, the notion of which cells represent the suburbs or the city center is replaced by how many cells have the highest concentration. Therefore, while the specific cells with the highest population may randomly change, the number of cells with a high population may be invariant.

Random field theory is often used for brain imaging [12] to identify regions of high activity. Inspired from this, the key parameter in our model is based on the number of cells with high population concentrations consuming large amounts of energy. The advantage of adopting this approach for a metropolis is that it reflects the dynamic traffic and energy consumption patterns without being tied to specific mobility models. Under our approach, the population randomly moves across any cell of the metropolis. This allows us to determine the energy statistics when the population has maximal mobility and contrast them against when mobility patterns are constrained to smaller areas. Our approach thus sheds light on arbitrary population mobility models.

2. We derive scaling laws on the endogenous growth of population and the metropolis area (as counted by the number of cells) in terms of two complementary inequalities. In the first law, namely the *cell growth*, the population is fixed, but the number of cells (or area) increases until the metropolis population density drops below a certain threshold. We show that this threshold corresponds to when the exterior area adjacent to the city's boundary equals the average unoccupied area in its interior. In the second approach, namely the *population growth*, the number of cells is constant, but the population increases only if a certain minimum fraction of the metropolis area is neither underpopulated nor overpopulated. We define two simple inequalities that denote the Pareto-optimal equilibria, which determine whether only the population or only the metropolis area can increase.

3. Given a constant metropolis area, we evaluate the mean and the variance of the consumed energy statistics as a function of population. As the population increases for a constant metropolis area, the variance and entropy in consumed energy eventually diminish to zero as the corresponding mean consumed energy approaches an asymptotic limit. With initial increases in the population, there is increasing variation in consumed energy. However, after a certain point, the population concentration for most cells becomes too high, leading to saturation in consumed energy. We show that the variance of consumed energy is bounded by a quadratic function of the city's mean consumed energy under certain circumstances. If the consumed energy is

representative of a city's economic size, then we also show that given a constant population density, as the city's area grows, its domestic product scales linearly with this area.

4. We show that the variance (and entropy) of the consumed energy decline when the population is proportionally divided into smaller, discontinuous and isolated subregions rather than being one whole. This implies that the variance or entropy represents the degree of mobility (spatial freedom) of the population.

5. From an information theoretic perspective, the mutual information of the consumed energy between any given subregion and the remaining area of the metropolis diminishes if the areas of the two zones are disproportional. If consumed energy is a measure of the economy, it means that smaller subregions are a less representative sample of the economic characteristics of the whole city. Our results reinforce the results in [5] that the marginal utility of information declines with subscaling.

6. We validate aspects of our model using power consumption data from Singapore. Specifically, we empirically show that as the mean of energy consumption crosses a point, the variance begins to diminish, provided that the city's area remains constant.

The structure of this paper is as follows. In Section 2, we present a system model to represent a metropolis. In Section 3, we then postulate scaling laws and equilibrium points for the growth of population and metropolis spatial spread. The second-order statistics (i.e., mean and variance) of the consumed energy are analyzed in Section 4. We then explore the entropy properties of the consumed energy and present results on the mutual information between subscales of the metropolis in Section 5. In Section 6, we provide empirical evidence of our model using power-demand data from Singapore. Finally, Section 7 provides conclusions and directions for future work.

2. System Model

Let a metropolis be partitioned into s identical geographical cells, each of unit area, representing resource cells such as land, housing or electric grid infrastructure. The partitioning represents an equal division of the metropolis area into cells, where each may consume the same maximum energy. The cells are located in a pattern that may be square, circular or some other regular shape. The region is occupied by n population units, where each unit may be an individual person or a large grouping of persons. Without loss of generality, we assume that each of n population units can randomly move between any of the s cells with uniform distribution in successive units of time (or iterations). A cell may be occupied by zero, one or more population units

in any iteration. A uniform distribution for mobility represents the case when the population units have maximal mobility and may migrate between all cells freely. We assume *ergodicity*, where the iterated ensemble statistics denote the *short-term* temporal population movements.

We let $L_j : L_j \leq s$ be a random variable that denotes the number of cells occupied by j population units. For instance, $L_0 = 80$, $L_1 = 100$ and $L_2 = 50$ indicates that 80 cells have no population units, 100 cells are each occupied by a single population unit, 50 cells are each occupied by two population units and so on. To reiterate, a population unit may comprise an arbitrarily large set of individuals. Hence, L_0 essentially represents the number of quasi-empty or unoccupied cells with little or no population. In Figure 1, for example, the total number of deep blue cells equals L_0 .

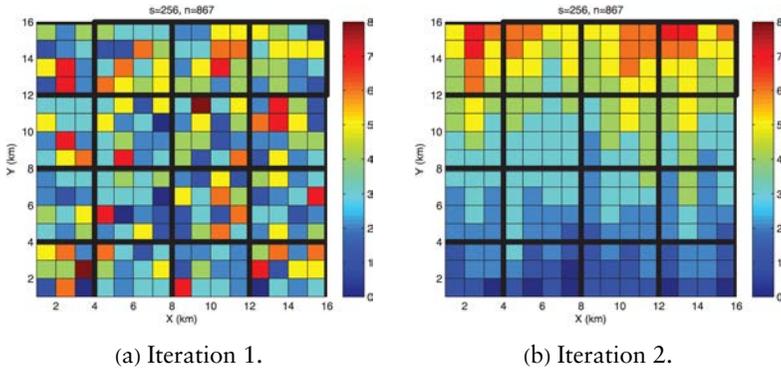


Figure 1. In an epoch, the population randomly moves between cells in the contiguous two-dimensional metropolis region with $k = 16$ subregions. A deep blue cell has no population unit.

When both n and s are large, a Poisson random distribution with intensity parameter (i.e., metropolis population density)

$$\lambda = \frac{n}{s}, \tag{1}$$

is a good approach to determine the average number of population units occupying the cells (as validated by Monte Carlo simulation). The average number of cells occupied by j population units is denoted as

$$\mathbb{E}[L_j] = s \frac{\lambda^j}{j!} \exp(-\lambda) = s \frac{(n/s)^j}{j!} \exp\left(-\frac{n}{s}\right), \tag{2}$$

where $E[\bullet]$ is the expectation operator. To reiterate, while the population units randomly move back and forth between the s cells, $E[L_j]$ denotes the average number of cells that will each be occupied by j population units.

2.1 Consumed Energy

The energy consumed by each of the s cells comprises an arbitrary mix of solar, fossil fuel and alternative sources. This hybrid mix is presumed identical for each cell and is calibrated to scale between a minimum of E_{\min} and a maximum of E_{\max} units of energy. The lower limit is due to an irreducible energy consumption per cell, whereas the upper limit is due to a constraint on how much energy a cell may maximally access. For instance, if a cell represents a housing unit, then even with full occupancy, the consumed energy of such a unit will not exceed some maximum supplied energy. Likewise, the units will always require some minimum energy for maintenance or sustenance. Between these two limits, however, the consumed energy depends on the number of occupying population units. In an iteration, a simple energy-consumption model for each cell $i \in \{1, 2, \dots, s\}$ as a function of the number of population units j occupying it, is as follows:

$$M_i = \begin{cases} \frac{E_{\max} - E_{\min}}{t}j + E_{\min}, & \text{if } j < t \\ E_{\max}, & \text{if } j \geq t, \end{cases} \quad (3)$$

where t is a saturation threshold. The consumed energy \mathcal{E} of the metropolis in any given iteration is

$$\mathcal{E} = \sum_{i=1}^s M_i = \left(\sum_{j=0}^{t-1} L_j \left(\frac{E_{\max} - E_{\min}}{t}j + E_{\min} \right) + \sum_{j=t}^s L_j E_{\max} \right). \quad (4)$$

As the population moves about, the random variable \mathcal{E} changes accordingly. A depiction of the energy-consumption variations is shown in Figure 2 for the population distribution in Figure 1(a). A higher saturation t threshold leads to greater variation in the energy-consumption distribution.

2.2 Contiguous Region or Discontiguous (Isolated) Subregions

The metropolis can be viewed as comprising $k: 1 \leq k \leq s$ subregions, where each subregion comprises a fixed set of s/k cells. For tractability, we assume that k is such that s/k is a natural number. If these subregions are part of a contiguous region, all the n population units can move between all the s cells. Conversely, if all these k subregions are discontiguous, then the population in each is essentially isolated

from the others. In this latter case, we presume that each subregion then has n/k population units also, where the randomness in population mobility is independent and identically distributed (iid) with uniform distribution within each subregion of s/k cells.

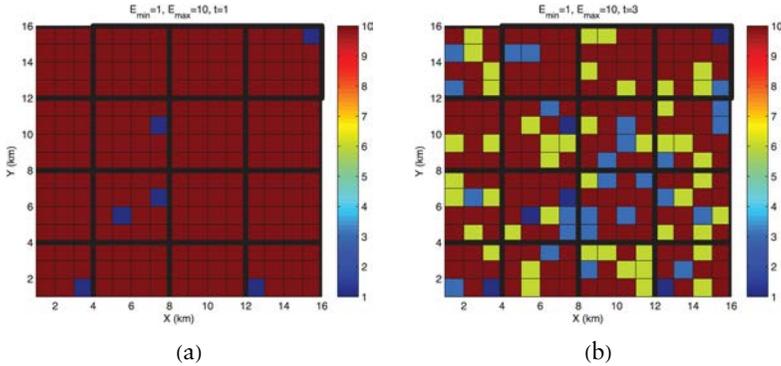


Figure 2. Given (a) two and (b) four possible quantized energy levels per cell, the equivalent heat map of the energy consumption of the metropolis for the population distribution in Figure 1(a).

3. Scaling Laws

In our proposed model, the scaling in both the surface area and the population of the metropolis occurs over epochs of much longer time duration than the short-term iterations. We let the metropolis be characterized by its area and population in a given *epoch* by (n, s) , where the population units follow the short-term iterative behavior outlined earlier. In the next epoch, the pair (n, s) representing the metropolis may evolve to different values.

We denote the average number of unoccupied (or quasi-empty) cells as $A_e = E[L_0]$ and the average number of occupied cells as $A_o = s - A_e$, respectively. Using equation (2), these variables equal

$$A_e = s \frac{(n/s)^0}{0!} \exp\left(-\frac{n}{s}\right) = s \exp\left(-\frac{n}{s}\right) \tag{5}$$

$$A_o = s - A_e = s \left(1 - \exp\left(-\frac{n}{s}\right)\right). \tag{6}$$

3.1 Sparsity Threshold

Consider the situation when the number of population units is such that the average number of occupied cells equals the average number

of unoccupied cells. Equivalently, we have

$$\begin{aligned} A_e &= A_o \\ n &= s \log(2). \end{aligned} \quad (7)$$

The relationship in equation (7) can be generalized to when the average number of unoccupied cells is more than s/k cells,

$$\begin{aligned} A_e &\geq \frac{k}{s} \\ n &\leq s \log(k), \end{aligned} \quad (8)$$

where $1 \leq k \leq s$. Given the pair (n, s) , if the inequality in equation (8) is satisfied, then the metropolis is deemed sparsely populated. Specifically, equation (7) represents the sparsity threshold of the metropolis.

■ 3.2 Cell Growth Equilibrium

We consider the situation when the population is fixed but the metropolis area, as denoted by the number of cells s , can increase in successive epochs. Under this approach, in each epoch, there is expansion of the metropolis area, provided that the population is sufficiently large.

The increase in the number of cells occurs along the edge of the metropolis area and the scaling of the area is therefore proportional to the metropolis's perimeter. Given a constant population n , we postulate that the expansion stops when the area gained through new cells along the perimeter equals the average unoccupied area within the metropolis. Throughout the expansion, all new cells remain contiguous to the city.

Let the scaling factor in the number of cells be $k : k > 1$, where s increases to ks . The unoccupied metropolis area must not become more than the area of the cells along the perimeter

$$\begin{aligned} A_e &\leq ks - s \\ sk \exp\left(-\frac{n}{sk}\right) &\leq s\left(1 - \frac{1}{k}\right) \\ n &\geq sk \log\left(1 - \frac{1}{k}\right). \end{aligned} \quad (9)$$

Thus, there is spatial growth of the metropolis area by a factor of $(1 - (1/k))$. For a square region, such as that shown in Figure 1, each of the s cells is unit area (1×1). For example, a city in the square grid will expand from 16×16 in successive epochs to 17×17 , 18×18 , 19×19 and finally to a 20×20 square region, as shown in Figure 3(a).

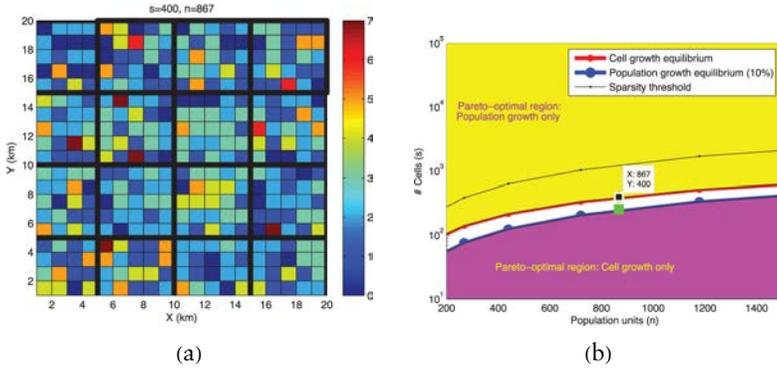


Figure 3. (a) The metropolis area grows from 256 to 400 cells, where the area adjacent to its perimeter equals the average unoccupied area (deep blue cells) within. (b) The equilibria relationship between population and cells is depicted here.

Hence, the increase in the number of cells at the perimeter in one scaling increment along each axis is about $4(\sqrt{s} + 1)$. Specifically, here $k = 1 + (4 / \sqrt{s}) + (4 / s)$, and the relationship in equation (9) is then

$$\begin{aligned}
 n &\geq sk \log(1 - 1 / k) \\
 n &\geq s \left(1 + \frac{4}{\sqrt{s}} + \frac{4}{s} \right) \left(\log(4 + \sqrt{s}) - \log(4) \right).
 \end{aligned}
 \tag{10}$$

The shape of the metropolis region can affect the expression in equation (10). For instance, with a circular metropolis region, the increase in the number of cells in one scaling increment in each direction by one along the circumference is approximately $2\sqrt{\pi s}$ with $k \approx 1 + 2\sqrt{\pi / s}$. This modifies the expression to

$$n \geq s \left(1 + 2\sqrt{\pi / s} \right) \left(\log(2\sqrt{\pi} + \sqrt{s}) - \log(\sqrt{\pi}) \right)$$

correspondingly.

Moreover, for an advanced civilization that exists on a theoretical Dyson’s sphere [13], the surface area around the star is $4\pi R^2$, where R is the spatial radius from s . As the civilization or the mega-structure metropolis grows on the shell encircling the star, its radius R increases and the cells expand on the two-dimensional spherical surface. To reiterate, the number of cells, all of unit area, grows proportionally to the surface area. With simple manipulation, we obtain

$$n \geq s \left(1 + \frac{4\pi}{\sqrt{s}} \right) \left(\log(4\pi + \sqrt{s}) - \log(4\pi) \right).
 \tag{11}$$

In all three metropolis shapes that span two-dimensional surfaces, we can deduce that the metropolis surface area will continue to grow unless the following happens:

$$n \leq \frac{s}{2} \log(s). \quad (12)$$

We have a cell growth equilibrium if the preceding inequality is satisfied. The area along the metropolis periphery then is smaller than the average empty area within. An equivalent interpretation of this equilibrium is that the metropolis perimeter becomes less than or equal to the maximal average perimeter of unoccupied cells inside the city. For instance, with square metropolises, the former is $4\sqrt{s}$, whereas the latter can be no more than $4s \exp(-n/s)$ on average. The relationship between these perimeters can be shown to also reduce to $n \leq (s/2) \log(s)$. *The area growth toward the exterior cannot happen, since there is too much unoccupied area within the interior.*

■ 3.3 Population Growth Equilibrium

Next we consider population growth given a constant metropolis area between successive epochs. As in Conway's *cellular automaton* [14], we postulate that the population will only increase if there are enough cells that are *critically populated*: cells that are neither underpopulated nor overpopulated. Between successive epochs, given a fixed s , the population n may increase to any larger value as long as there will be some minimum number of cells on average that will be critically populated.

We define the average number of underpopulated cells as $\mathcal{E}[L_0]$ and overpopulated cells as $\mathcal{E}[L_j] : j \geq 2$. Thus, the average number of critically populated cells are ones with a single population unit (i.e., $\mathcal{E}[L_1]$). While other thresholds may also be calibrated for critical population, this serves as a useful starting point. The average number of critically populated cells is then

$$A_1 = s \frac{(n/s)^1}{1!} \exp\left(-\frac{n}{s}\right) = n \exp\left(-\frac{n}{s}\right), \quad (13)$$

which is the mean number of cells with a single population unit. Increases in the metropolis population will only occur if this is sufficiently large. The maximum possible increase in n between successive epochs can be generalized such that the average number of critically populated cells is at least s/k . *In other words, $1/k$ of the metropolis area on average must be neither underpopulated nor overpopulated to support further population growth.* When this is not satisfied, we have $\mathcal{E}[L_1] \leq s/k$, where $1 \leq k \leq s$,

$$\begin{aligned}
 A_1 &\leq \frac{s}{k} \\
 \exp\left(-\frac{n}{s}\right) &\leq \frac{s}{k} \\
 n &\leq s\left(\log\left(\frac{n}{s}\right) + \log k\right).
 \end{aligned}
 \tag{14}$$

Population growth equilibrium will be reached if this inequality is satisfied, where the average number of critically populated cells falls below $1/k$ of the metropolis area. At minimum, for population growth, there must always be at least one cell on average that is critically populated, where the alternative to equation (14) simplifies to

$$n \geq s \log(n). \tag{15}$$

■ 3.4 Growth Cycles

Cell growth and population growth equilibria may occur in successive epochs. For instance, starting from some $(n^{(1)}, s^{(1)})$ in an epoch, there may be cell growth to $s^{(2)}$ until equilibrium is reached (i.e., $n^{(1)} \leq (s^{(2)}/2) \log(s^{(2)})$). In the next epoch, given the updated state $(n^{(1)}, s^{(2)})$, the population grows until we achieve $n^{(2)} \leq s^{(2)} \log(n^{(2)})$. Thus, the cycles may alternate between cell growth and population growth over the epochs.

We illustrate the concept of metropolis area (cell) growth given a high-enough population density. In Figure 1, we depict two population distribution ensembles of a metropolis with $s = 256$ cells. With $n = 867$ population units that randomly move between the cells, the population density is $\lambda = n/s \approx 3.4$, with an average of nine unoccupied as per equation (10). Also note that as per equation (14), the pair $(867, 256)$ represents a population growth equilibrium for $k = 10$ or 10% area threshold.

Given a constant n , the area scales proportionally to the perimeter until we have $s = 400$, as shown in Figure 3(a). We then have $\lambda \approx 2.2$ population units, with an average of 46 unoccupied cells. At this point, any further cell growth by one cell along the city's perimeter would add 40 new cells, which is lower than the average number of unoccupied cells (i.e., 46). The cell growth reaches equilibrium at this point.

In reality, both cell and population growth may happen simultaneously. The inequalities (10) and (14) therefore represent Pareto-optimal boundaries between cell growth and population growth equilibria, where only one or the other happens. In Figure 3(b), we plot the relationship between endogenous growth of population and a

square-shaped metropolis area by setting (10) and (14) to equations. We obtain pairs of (n, s) that satisfy the equations

$$n = s \left(1 + \left(4 / \sqrt{s} \right) + (4 / s) \right) \left(\log(4 + \sqrt{s}) - \log(4) \right)$$

and

$$n = s(\log(n/s) + \log k),$$

respectively, through a numerical method. Note that any coordinate (n, s) that lies above the former equation represents a point that cannot experience cell growth, as the population density is too low. Conversely, if (n, s) lies below the latter equation, we cannot have population increase, as there are too few critically populated cells.

Figure 3(b) is a log-linear plot that shows the narrow region (shaded white) between the two equations where both population and cells can increase ($k/s = 0.1$ or 10% of the metropolis area). The transition between two equilibria is shown for the coordinates denoting the pairs of (n, s) for two distinct epochs. The line labeled *sparsity threshold* represents the relationship in equation (7), where the average number of occupied cells equals the average number of unoccupied cells.

The average number of critically populated cells, which are deemed neither overpopulated nor underpopulated, determines how much increase in n is possible. In the log-log plots in Figure 4(a), for population growth equilibrium, the threshold for critically populated cells is set to 10% of s as well as to just a single cell. When n is large enough, the population growth Pareto-optimal equation becomes the largest for 10%. Conversely, if a single cell is needed to be critically populated on average, the equation has a lower gradient and does not intersect the cell growth equilibrium line.

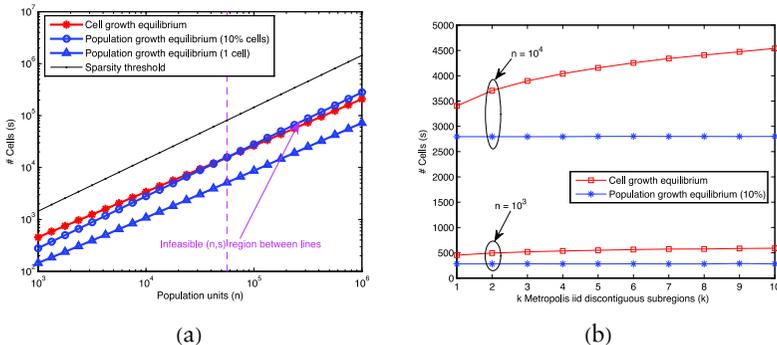


Figure 4. (a) A contrast between equilibria points. (b) With discontinuous subregions, there is additional cell growth due to their larger aggregate perimeters as compared to a contiguous region. The population growth equilibrium is invariant to this factor.

There is also another interesting implication for the coordinates in Figure 4(a) for large populations (i.e., $n > 10^{4.75}$) when the line for the population growth equilibrium at 10% is higher than that for the cell growth. Sandwiched between the two lines, these coordinates represent an infeasible set of pairs (n, s) . The coordinates below the former line denote overcrowded population, and coordinates above the latter line indicate too much unoccupied area within the metropolis for cell growth. Hence, for these values, neither kind of growth is possible. Conversely, if the population growth threshold is based on a fixed number of critically populated cells, then this restriction does not apply.

3.5 Discontiguous Subregion Growth

In Figure 4(b), we plot the relationship between n and s assuming an equality constraint (i.e., exact equilibria) set in equations (10) and (14) for a square metropolis area, assuming that the city comprises κ isolated and discontiguous subregions. In this case, the variable $\kappa: 1 \leq \kappa \leq s$ represents a subscaling factor denoting the number of smaller discontiguous subregions, each with population n/κ and with s/κ cells self-similarly organized as square subregions. Each of these subregions will exhibit the same growth cycles between epochs as described earlier. We presume that none of these κ subregions will physically overlap as their areas grow and will always remain discontiguous. Modifying equation (10) for the case where we have

$$\begin{aligned} \frac{n}{\kappa} &\geq k \left(\frac{s}{\kappa} \log(1 - 1/k) \right) \\ n &\geq \left(s \left(1 + \frac{4}{\sqrt{s/\kappa}} + \frac{4}{s/\kappa} \right) \left(\log(4 + \sqrt{s/\kappa}) - \log(4) \right) \right). \end{aligned} \tag{16}$$

Conversely, with κ isolated subregions for the constraint on population growth, we observe that modifying equation (14) yields

$$\begin{aligned} \frac{n}{\kappa} \left(\exp\left(-\frac{n/\kappa}{s/\kappa}\right) \right) &\geq \left(\frac{s/\kappa}{k} \right) \\ n &\geq s \left(\log\left(\frac{n}{s}\right) + \log k \right). \end{aligned} \tag{17}$$

Figure 3(b) shows that with increasingly more subregions (or κ), the pairs (n, s) under cell growth equilibria have a higher aggregate s , where the population per subregion is at n/k . However, the converse is not true. The population growth equilibrium does not change. This means that when the subregions are isolated and discontiguous, there can be no extra population growth. However, with cell growth for each discontiguous subregion, the aggregate area gained by all subregions is more than if all these subregions were a part of a single

contiguous metropolis. To reinforce, this is due to the fact that the aggregate area of the perimeters of the k isolated subregions is more than that of a single contiguous region.

4. Energy Statistics

Next we derive the statistics of the consumed energy of the metropolises defined by the pair (n, s) in any given epoch. Recall that $\lambda = n/s$ is the population density. A closed-form expression of a probability mass function (pmf) of consumed energy is difficult to derive, owing to the number of variables involved. Hence, we simplify the problem by considering the mean and variance of the consumed energy of a single cell and scaling these metrics by the total number of cells to get a bound on the variance of the metropolis. Dividing equation (2) by s gives us the average fraction of the time a given cell is occupied by j population units. Using this approach, the mean consumed energy of the metropolis is

$$E[\mathcal{E}] = s \left(\sum_{j=0}^t \left(\frac{E_{\max} - E_{\min}}{t} j \right) \frac{\lambda^j \exp(-\lambda)}{j!} + E_{\max} \left(\sum_{j=0}^t 1 - \frac{\lambda^j \exp(-\lambda)}{j!} \right) \right), \tag{18}$$

where $E[L_o] = s \exp(-\lambda)$. The corresponding variance in the consumed energy of the metropolis is

$$\text{Var}(\mathcal{E}) = E[\mathcal{E}^2] - (E[\mathcal{E}])^2. \tag{19}$$

This can be expanded as

$$\text{Var}(\mathcal{E}) = E \left[\left(\sum_{j=0}^{t-1} L_j \left(\frac{E_{\max} - E_{\min}}{t} j + E_{\min} \right) + \sum_{j=t}^s L_j E_{\max} \right)^2 \right] - (E[\mathcal{E}])^2. \tag{20}$$

The variance of the metropolis consumed energy is constrained by the variance of the consumed energy of any given cell multiplied by the total number of cells s as follows:

$$\text{Var}(\mathcal{E}) \leq s \times \underbrace{\left(\sum_{j=0}^t \left(\frac{E_{\max} - E_{\min}}{t} j \right)^2 \frac{\lambda^j \exp(-\lambda)}{j!} + E_{\max}^2 \left(\sum_{j=1}^t 1 - \frac{\lambda^j \exp(-\lambda)}{j!} \right) \right)}_{\text{second moment cell energy}}$$

$$- \left[\underbrace{\left(\sum_{j=0}^t \left(\frac{E_{\max} - E_{\min}}{t} j \right) \frac{\lambda^j \exp(-\lambda)}{j!} + E_{\max} \left(\sum_{j=0}^t 1 - \frac{\lambda^j \exp(-\lambda)}{j!} \right) \right)}_{\text{first moment cell energy}} \right]^2. \tag{21}$$

For $t = 1$, we obtain a simple expression for the mean in equation (18) as

$$E[\mathcal{E}] = s((E_{\min} - E_{\max}) \exp(-\lambda) + E_{\max}) \tag{22}$$

and the bound on the variance in equation (20) as

$$\begin{aligned} \text{Var}(\mathcal{E}) &\leq s[(E_{\min}^2 - E_{\max}^2) \exp(-\lambda) + E_{\max}^2] - \\ &\quad \left[(E_{\min} - E_{\max})^2 \exp(-2\lambda) - \right. \\ &\quad \left. 2E_{\max}(E_{\min} - E_{\max}) \exp(-\lambda) + E_{\max}^2 \right] \\ &\leq s[(E_{\min}^2 + E_{\max}^2) \exp(-\lambda) - \\ &\quad (E_{\min}^2 + E_{\max}^2) \exp(-2\lambda) - 2E_{\min}E_{\max} \exp(-\lambda)] \end{aligned} \tag{23}$$

$$\text{Var}(\mathcal{E}) \leq s(\exp(-\lambda) - \exp(-2\lambda)) (E_{\min} - E_{\max})^2.$$

These equations represent elegant relationships connecting the variance of consumed energy to the number of cells in the metropolis, its population density and the minimum and maximum energy per cell. At $t = 1$, the mean consumed energy and the bound on the variance in equations (22) and (23) are combined as follows:

$$\begin{aligned} \exp(-\lambda) &= \frac{\frac{E[\mathcal{E}]}{s} - E_{\max}}{E_{\min} - E_{\max}} \\ \text{Var}(\mathcal{E}) &\leq s \left[\frac{E[\mathcal{E}]/s - E_{\max}}{E_{\min} - E_{\max}} \right] \left[1 - \frac{E[\mathcal{E}]/s - E_{\max}}{E_{\min} - E_{\max}} \right] (E_{\min} - E_{\max})^2. \end{aligned} \tag{24}$$

This leads to a straightforward quadratic relationship between the mean consumed energy and the variance of the metropolis as

$$\text{Var}(\mathcal{E}) \leq \frac{(E[\mathcal{E}] - sE_{\max})(sE_{\min} - E[\mathcal{E}])}{s}. \tag{25}$$

In subsequent results, we use the Monte Carlo method to generate the statistics for consumed energy. By generating a large number of random samples with n population units uniformly placed in the s cells, the method provides the desired energy statistics. The energy parameters set in our Monte Carlo trials are $E_{\min} = 1$ and $E_{\max} = 10$ units of energy. These values are chosen to reflect that the maximum energy any resource cell consumes may be an order of magnitude higher than the minimum.

In Figure 5, we plot the mean and the variance of the consumed energy using the Monte Carlo method for various pairs of (n, s) . For a

fixed metropolis area, as denoted by a constant s , we evaluate these statistics for various population sizes of n ranging from 100 units to 2600 units. We also have two values of the saturation threshold t for illustration. In Figure 5(a), we observe that for small values of n , the consumed energy scales linearly before asymptotically converging to the peak threshold of sE_{\max} . Thus, with increasing n we have mean consumed energy as $E[\mathcal{E}] = 4000$ for $s = 400$ and $E[\mathcal{E}] = 2000$ for $s = 200$. The scaling of variance in the consumed energy, however, exhibits a concave shape that diminishes to zero for large n for both values of s in Figure 5(b). A higher saturation threshold at $t = 3$ naturally tends to shift the curves, where the mean and variance of the consumed energy in both Figure 5(a) and Figure 5(b) reach their maximum at higher values of n .

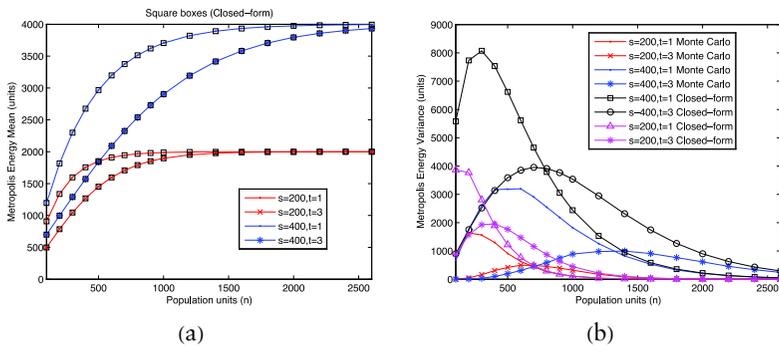


Figure 5. Given s , the mean metropolis energy converges to the maximum for larger values of n in (a), whereas in (b), the corresponding variance in the consumed energy first peaks and then gradually declines to zero.

In Figure 5(a), the mean derived in equation (18), represented by square coordinates, perfectly matches the mean consumed energy obtained through the Monte Carlo method. In Figure 5(b), the inequality in equation (20) illustrates the bound on the variance of consumed energy.

Hence, we observe non-monotonicity and concavity of the variance of the consumed energy for fixed s over the range of values of n . In other words, if the metropolis area is constant, smaller populations have less randomness in the consumed energy. With much larger populations, there will likely be multiple population units occupying a majority of cells. In this latter case, the consumed energy saturates for most cells, leading to a decline in the variance.

4.1 Consumed Energy and Economic Size

The size of the metropolis economy can also be indicated by its overall consumed energy [9, 10]. Thus, the metropolis domestic product

(MDP) can be hypothesized to be linearly proportional to the latter. Let $E[\mathcal{E}^{(1)}]$ and $E[\mathcal{E}^{(2)}]$ denote the mean consumed energy in two different epochs. Allowing M_1 and M_2 to then denote the magnitudes of their MDPs, respectively, we may state that

$$\frac{M_2}{M_1} = \frac{E[\mathcal{E}^{(2)}]}{E[\mathcal{E}^{(1)}]}. \tag{26}$$

Suppose that the population density $\lambda = n^{(1)} / s^{(1)} = n^{(2)} / s^{(2)}$ is constant between the two epochs and that there is growth in the metropolis area by a factor of $k : k \geq 1$ (i.e., s to ks). We surmise that $E[L_j] = ks(\lambda^j / j!) \exp(-\lambda)$. In this case, by modifying equation (4), we have

$$\mathcal{E}^{(2)} = \left(\sum_{j=0}^{t-1} kL_j \left(\frac{E_{\max} - E_{\min}}{t} \cdot j + E_{\min} \right) + \sum_{j=t}^{s^{(2)}} kL_j E_{\max} \right) = k\mathcal{E}^{(1)}. \tag{27}$$

Thus, with constant population density, as the number of resource cells increases from s to ks , equation (26) simplifies to

$$\frac{M_2}{M_1} = k \frac{E[\mathcal{E}^{(1)}]}{E[\mathcal{E}^{(1)}]} = k. \tag{28}$$

In other words, given a constant population density, the average energy consumed scales linearly with s . *Thus, the growth in MDP is then simply proportional to the overall increase in the mean consumed energy or equivalently, the gain in the metropolis area.*

5. System Entropy

The energy consumed by each of the s cells can be in $t + 1$ energy states. Let the pmf of the consumed energy of the metropolis area be defined by $p_{\mathcal{E}}(z)$, where z denotes the z^{th} state from among the s^{t+1} possible discrete or quantized consumed energy levels for the entire metropolis. The pmf can also be obtained through the Monte Carlo method.

The Shannon entropy [11] of the consumed energy is a function of variables n , t and s such that

$$H(n, s) = - \sum_{z=1}^{s^{t+1}} p_{\mathcal{E}}(z) \cdot \log(p_{\mathcal{E}}(z)). \tag{29}$$

Mutual information between any two subregions represents how similar the two zones are in terms of consumed energy. The work in [5]

provides an interesting exposition, using entropy to observe the dependencies of various subcomponents of a large system.

In this context, we consider the metropolis comprising k subregions of equal area. If these subregions are discontinuous and isolated, with n/k population units (iid), the consumed energy entropy of each subregion is notated as $H(n/k, s/k)$. The entropy of the metropolis is then simply the sum of the entropies of all subregions

$$H(n, s) = \sum_i^k H\left(\frac{n}{k}, \frac{s}{k}\right). \quad (30)$$

However, when the k subregions are part of one contiguous region, where all n population units can move to any cell, then the entropy of each subregion is denoted as $H(n, s_1)$. The consumed energy of a subregion is correlated to the consumed energy of the remaining subregions, owing to greater population mobility. The mutual information between any one of the subregions and the other remaining $k - 1$ subregions in the contiguous region is defined as

$$I(s_1, s_{k-1}) = H(n, s_1) + H(n, s_{k-1}) - H(n, s). \quad (31)$$

Note that $H(n, s_k)$ is the entropy of subregion 1 (identical for all k subregions) and $H(n, s_{k-1})$ is the entropy of the set of the $k - 1$ remaining subregions. Note that in the discontinuous case, the mutual information between any two subregions is always zero.

In Figure 6(a), we plot the pmf of the consumed energy for three values of n . We can observe that the spread of the pmf, which essentially corresponds to the variance in consumed energy: (1) is small for $n = 100$; (2) increases for the intermediate case $n = 400$; and (3) eventually diminishes again for $n = 2000$. In Figure 6(b), we plot the entropy equation (29) of the consumed energy over the range of n for two thresholds of both s and t . We observe that the maximum entropy occurs at the same values of n where the variance in the consumed energy became maximum in Figure 5(b).

In Figure 7, we plot the variance and entropy of a single subregion both in the contiguous and discontinuous settings. Note that $k = 1$ corresponds to the default case with one subregion. We can see that with increasing k , these metrics decline quickly in both situations. As the population units are free to migrate across all regions in the contiguous setting, we observe a higher variance and entropy as compared to when the population is isolated into discontinuous pockets. *It implies that variance and entropy of the metropolis consumed energy indicate the spatial mobility or spatial freedom of the population. Arbitrary mobility patterns of the population would thus only produce a lower variance or entropy as compared with when the population can move between all cells.*

Figure 8 plots the mutual information between any single subregion and the remaining $k - 1$ subregions for different values of k . Essentially, we illustrate how representative a given subregion is of the remaining metropolis area. When $k = 2$, the metropolis has two subregions, where naturally the mutual information between them is also high. When $k = 100$, the mutual information diminishes significantly for any given n , as any subregion is one one-hundredth of the metropolis area and provides a poor approximation of the consumed energy pattern for the remaining 99 subregions. Our results also reinforce the results in [5] that the marginal utility of information declines with scale.

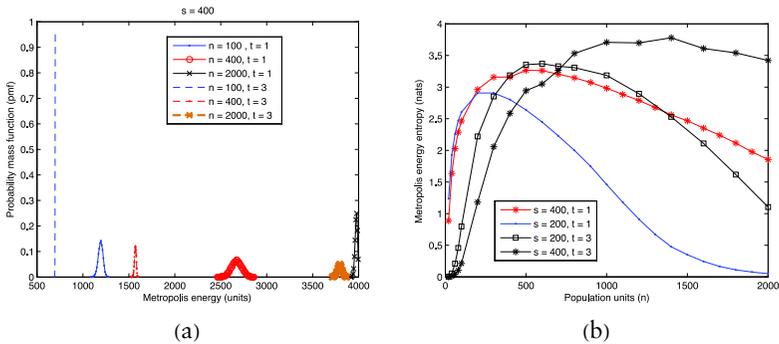


Figure 6. (a) The pmf of the consumed energy with various population thresholds. (b) The entropy of the metropolis shows a trend similar to the variance in Figure 5(b).

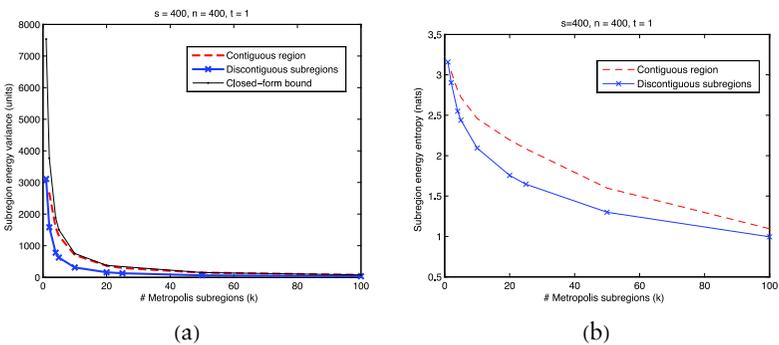


Figure 7. The (a) variance and (b) entropy of any single subregion of the metropolis. With discontiguous subregions, these metrics are lower than when there is a contiguous metropolis.

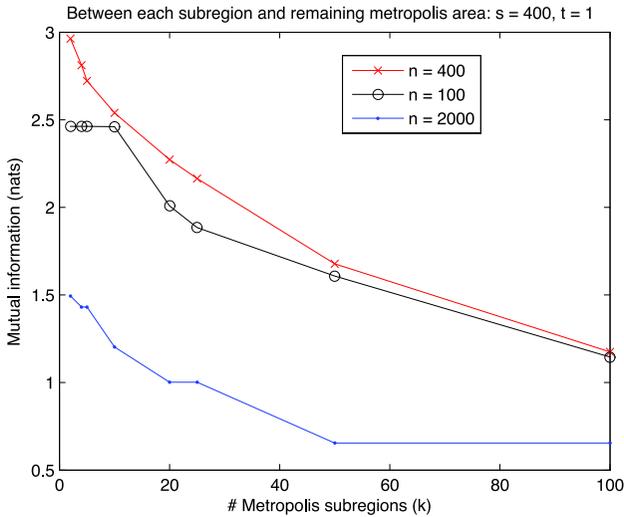


Figure 8. The mutual information between any one subregion and the remaining $k - 1$ subregions declines with k . When $k = 2$, we essentially have maximum similarity of the consumed energy patterns of the two large metropolis subregions.

6. Empirical Analysis

We use data from the Singapore power grid to examine the impact of urbanization on energy consumption to validate aspects of our model. Singapore is a good candidate metropolis to study the impact of urbanization on energy consumption. We also performed a preliminary search of the power-demand data from some other nations, including the US. However, there are some limitations of that data, where, for example, the hourly data from the US Energy Information Administration either only goes back to July 2015 or is not filtered for specific metropolitan zones (e.g., Chicago or Los Angeles). In contrast, using the Singapore Energy Statistics dataset [15], we have system power demand (in megawatts) sampled at every half-hour interval from 2004. The power demand is our measure of the iterating energy consumption, as modeled earlier. Additionally, Singapore's area is about 279 square miles with a large urban area (at least 75% of land mass) [16]. Given that the city-state is a well-developed island of a fixed area, urbanization is mostly driven by population growth.

We focus our attention on the samples obtained for all days from the first Monday in February until the first Sunday in March. This period coincides with the tropical city's minimal average rainfall [17]. Our choice of this annual period is to reduce environmental factors

that can otherwise affect or disrupt power demand. The subsequent analysis is based on 20 000 sample points over the 15-year period.

To reiterate, since the urban area is relatively constant, the long-term population growth is a key driver of increase in the power demand. In Figure 9(a), we plot the impact of population growth on the average power demand. The latter is the mean of all power-demand samples collected in the February–March interval from each year. As expected, we can generally see an increase in the average power demand with increasing population. In Figure 9(b), we then plot a scatter plot of the average power demand and standard deviations. We can clearly see that the average is the lowest in 2004 and increases to its highest value in 2019. However, the peak standard deviation for the 2004–2019 epoch occurs in 2013. The standard deviation (or variance) is also a measure of the entropy of the power demand. These results show that our basic metropolis energy consumption model has empirical grounding, where when the city size is smaller, the entropy increases with population, and when the size is larger, it diminishes. Nonetheless, more work is needed to validate the model by using data from other cities.

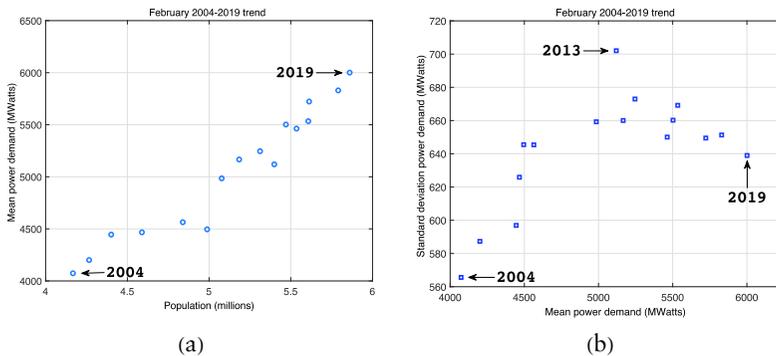


Figure 9. (a) Increasing population in Singapore coincides with increasing power demand; (b) the standard deviation (or variance) initially increases with mean power demand and then declines. These empirical results support the fundamental model presented in the paper.

In Figure 10, we plot two PDFs of the aggregated monthly power-demand samples (using 50-bin histograms) from 2004 and 2019. We observe a wide range of power-demand levels for both years, with two peaks at the extremes for each. These patterns do not neatly fit with well-known Gaussian or uniform distributions.

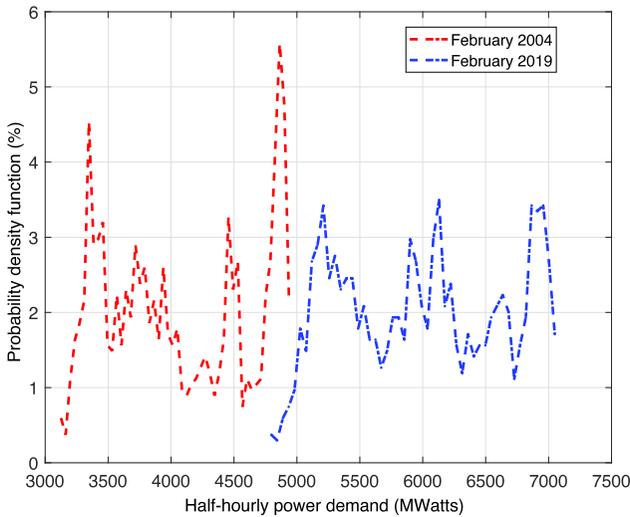


Figure 10. The distribution of Singapore's power demand interestingly does not follow a Gaussian distribution despite the several million consumers.

7. Conclusion

Complexity can be found in social networks and biological, engineering or economic systems. We have proposed a new approach to study the scaling laws, entropy and consumed energy of a metropolis that is a complex system comprising all these components. Using an abstract model of a metropolis in terms of two fundamental variables consisting of population and area (or cells, more broadly), we have shown that when the metropolis area is fixed, increasing population initially increments the consumed energy's variance or entropy up to a certain threshold, and then it declines afterward. Moreover, for the case when population and cells can mutually affect each other, we have derived equations that define the Pareto-optimal boundaries where either population increase or metropolis area expansion happens. We have also shown that variance or entropy of the consumed energy of a metropolis can represent the spatial freedom of a mobile population. Future work will consider applying these principles by calibrating them to additional data available for other metropolises in addition to Singapore. Our model may serve as a blueprint to examine a broad class of complex systems.

Appendix

A. PMF Derivation

Given the pair (n, s) , we derive the pmf of the number of cells occupied by at least one population unit through an iterative algorithm approach. As an illustration, consider a small system with $n = 4$ and $i = 2$ occupied cells. We iteratively put all four population units in the cells randomly, one by one, such that only $i = 2$ cells are occupied. In each iteration, let an event success S be defined as when a population unit is put in a cell that is currently unoccupied and failure F as when the population unit is put in a cell that already has a population unit. The very first population unit will always have success, as no cell is initially occupied. With additional population units randomly placed in the cells, the probability of failure will increase, as more and more of the cells are likely occupied. Specifically, for $i = 2$, we will have the following iteration combinations: $SSFF$, $FFFS$ and $SFSF$. Note that the corresponding probability of i occupied cells for each combination is

$$\frac{s}{s} \cdot \frac{s-1}{s} \cdot \frac{2}{s} \cdot \frac{2}{s}, \frac{s}{s} \cdot \frac{1}{s} \cdot \frac{1}{s} \cdot \frac{s-1}{s}$$

and

$$\frac{s}{s} \cdot \frac{1}{s} \cdot \frac{s-1}{s} \cdot \frac{2}{s},$$

respectively. We can infer that the probability of i successes is thus

$$\prod_{j=1}^{i-1} \left(1 - \frac{j}{s}\right).$$

However, the probability of failures depends on the prior set of occupied cells as population units are placed in them in succession. The first i population units should randomly go in any i unoccupied cells. The remaining $n - i$ population units then must fall in these i cells in successive iterations.

Since there are i successes, let $a_k : k \in \{1, 2, \dots, i\}$ be the number of consecutive failures between successive successes such that $SF^{a_1}SF^{a_2}\dots SF^{a_i}$ and where $\sum_{k=1}^i a_k = n - i$. We can deduce that the pmf that i cells are occupied by at least one population unit is

$$P_i = \prod_{j=1}^{i-1} \left(1 - \frac{j}{s}\right) \sum_{a_1=0}^{n-i} \sum_{a_2=0}^{n-i} \dots \sum_{a_i=0}^{n-i} \prod_{k=1}^i \left(\frac{k}{s}\right)^{a_m} : \sum_{m=1}^i a_m = n - i. \tag{A.1}$$

There are i summation terms in equation (A.1), with the constraint that the total number of failures equals $\sum_{k=1}^i a_k = n - i$. The computa-

tional complexity of obtaining the pmf of the number of occupied cells is prohibitive. As the pmf of the consumed energy depends on the distribution of occupied cells, a Monte Carlo method is a quicker and more efficient approach to determining $p_{\mathcal{E}}(z)$. In Figure A.1, we plot the pmf for various pairs of (n, s) and validate the expression through a Monte Carlo approach.

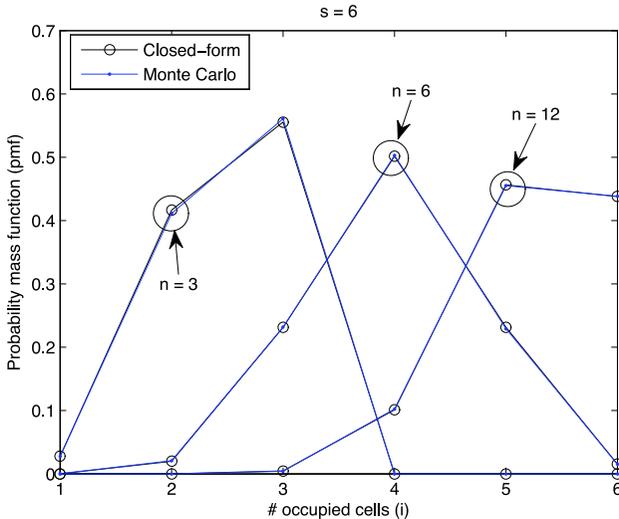


Figure A.1. The Monte Carlo approach validates the expression in equation (31), which is the pmf of the number of occupied cells where the latter is computationally intensive to obtain.

Acknowledgments

The author would like to express appreciation for the review and comments by Abdul Naseer Malmi Kakkada of the University of Texas at Austin.

References

- [1] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert and G. B. West, “Growth, Innovation, Scaling, and the Pace of Life in Cities,” *Proceedings of the National Academy of Sciences*, 104(17), 2007 pp. 7301–7306. doi:10.1073/pnas.0610172104.
- [2] G. Duranton and D. Puga, “The Growth of Cities,” *Handbook of Economic Growth*, Vol. 2 (P. Aghion and S. N. Durlauf, eds.), Amsterdam: Elsevier, 2014 pp. 781–853. doi:10.1016/B978-0-444-53540-5.00005-7.

- [3] J. H. von Thünen, *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie (The Isolated State)*, Hamburg: Wirtschaft & Finan, 1826.
- [4] Y. Bar-Yam. “General Features of Complex Systems.” New England Complex Systems Institute. (Aug 15, 2019) necsi.edu/s/Paper1.pdf.
- [5] B. Allen, B. C. Stacey and Y. Bar-Yam, “Multiscale Information Theory and the Marginal Utility of Information,” *Entropy*, **19**(6), 2017 273. doi:10.3390/e19060273.
- [6] P. Gupta and P. R. Kumar, “The Capacity of Wireless Networks,” *IEEE Transactions on Information Theory*, **46**(2), 2000 pp. 388–404. doi:10.1109/18.825799.
- [7] A. Conn, U. V. Pedmale, J. Chory, C. F. Stevens and S. Navlakha, “A Statistical Description of Plant Shoot Architecture,” *Current Biology*, **27**(14), 2017 P2078-2088.e3. doi:10.1016/j.cub.2017.06.009.
- [8] A. Elnakat, J. D. Gomez and N. Booth, “A Zip Code Study of Socioeconomic, Demographic, and Household Gendered Influence on the Residential Energy Sector,” *Energy Reports*, **2**, 2016 pp. 21–27. doi:10.1016/j.egyr.2016.01.003.
- [9] F. Karanfil and Y. Li, “Electricity Consumption and Economic Growth: Exploring Panel-Specific Differences,” Department of Research, IPAG Business School, 2014, 2014-337. (Aug 15, 2019) EconPapers.repec.org/RePEc:ipg:wpaper:2014-337.
- [10] E. Robinson, D. Egel, P. Johnston, S. Mann, A. D. Rothenberg and D. Stebbins, *When the Islamic State Comes to Town*, Santa Monica, CA: RAND Corporation, 2017. (Aug 15, 2019) www.rand.org/pubs/research_reports/RR1970.html.
- [11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Hoboken, NJ: Wiley-Interscience, 2006.
- [12] M. Brett, W. Penny and S. Kiebel, “Introduction to Random Field Theory,” *Human Brain Function*, 2nd ed. (R. S. J. Frackowiak, et al., eds.), Boston: Elsevier Academic Press, 2004 pp. 867–879. doi:10.1016/B978-012264841-0/50046-9.
- [13] F. J. Dyson, “Search for Artificial Stellar Sources of Infrared Radiation,” *Science*, **131**(3414), 1960 pp. 1667–1668. doi:10.1126/science.131.3414.1667.
- [14] M. Gardner, “Mathematical Games: The Fantastic Combinations of John H. Conway’s New Solitaire Game ‘Life’,” *Scientific American*, **223**(4), 1970 pp. 120–123. www.jstor.org/stable/24927642.
- [15] Energy Market Authority, “Singapore Energy Statistics.” (Nov 18, 2018) www.ema.gov.sg/Statistics.aspx.
- [16] The World Bank, “Singapore.” (Nov 18, 2018) data.worldbank.org/country/singapore?view=chart.
- [17] Meteorological Service Singapore, “Climate of Singapore.” (Nov 18, 2018) www.weather.gov.sg/climate-climate-of-singapore.