Article:

Solomon, G. and Wilson, A. (2026) Reducing dimensionality and complexity in fine-grain urban models. Complexity, 2026 (1). 6120118. ISSN: 1076-2787

https://doi.org/10.1155/cplx/6120118

Complexity

WILEY

RESEARCH ARTICLE OPEN ACCESS

Reducing Dimensionality and Complexity in Fine-Grain Urban Models

Guy Solomon[1] | Alan Wilson[2]

[1]School of History, Philosophy and Digital Humanities, University of Sheffield, Sheffield, UK | [2]Centre for Advanced Spatial Analysis, University College London, United Kingdom and the Alan Turing Institute, London, UK

Correspondence: Guy Solomon (g.solomon@sheffield.ac.uk)

Received: 28 October 2025 | **Revised:** 23 March 2026 | **Accepted:** 2 May 2026

Academic Editor: Ning Cai

ABSTRACT

A challenge in the ongoing development of mathematical and computer modelling of cities is to add detail, and this implies finer scales—space, sectors and time. This involves specifying high-dimensional arrays and hence large numbers of variables for comprehensive models. The objective of this paper is to show how such large numbers can be reduced by defining sets of principal flows, along with constructing tiers of zone systems, in such a way that computation can be handled at fine scales. The argument is illustrated with the retail model but can easily be extended to a variety of applications.

1 | The Challenge

Urban models need to be specified in terms of scale: spatial units (zones), sectors and time. Here, we investigate models which need a fine-grain spatial structure, and we tackle the problem of high dimensionality which generates interpretational complexity along with computational challenges. We aim to show that we can extract useful information from large matrices. We use the retail model with a two-dimensional matrix of variables to fix ideas. This paper provides the basis for comprehensive models founded on small-zone input output accounts which will be reported in a later paper. Multiple time scales would add a further dimension.

As the number of pairwise interactions within a zonal system increases quadratically (relative to the number of zones), fine-grain systems quickly become complex. Consider a geo-system which for England and Wales might be divided into the 35,672 lower layer super output areas (LSOAs). If we seek to model the whole system (where each LSOA may interact with every other LSOA), there would be approximately 35,672 interactions/flows—that is 1.27Bn. This number of equations, in a simple model, would have to be computed this number of times for each iteration [1].

Earlier work on this problem provides some useful connections, in particular offering tests on the effects of aggregation. Arbia and Petrarca [2], with an econometric approach, use measures of “mean flow” to test the impacts of various levels of aggregation in traditional gravity models. Batty and Sidkar [3], the first of a series of papers aggregate through increasing the size of zones and test the impact using “spatial entropy” as a measure of information loss. In neither of these cases are there formal linked tiers of spatial systems. Others, such as Hagen-Zanker and Jin [4], have relied on hierarchial clustering to reduce the number of zones. Roumpani [5] did develop a two-tier version of the Lowry [6] model, extending earlier work by Dearden and Wilson [7], Chapter 4. The distinctive feature of the approach offered in this paper is to define tiers along with tests of what are significant flows, and this leads to dramatic reductions in dimensionality. Formally, we are not engaged with aggregation: We simply define an appropriate model that recognises significant flows which can be calibrated in the usual way.

We consider in turn the retail model as an example (Section 2); how to construct subsets to represent principal flows and hence reduce dimensionality (Section 3); we then enumerate the model equations for the retail model (Section 4). In Section 5, we review the potential for reducing dimensionality, and in Section 6, we present a worked example. Next steps and future work are identified in Section 7.

2 | The Retail Example

We use a simple version of the standard retail model for illustrative purposes [8]. A definition of all symbols used is provided in the annex. For a single region with zones i, j, ... , from $1\,to N$[1]; let P_i be the population of zone i and e_i, the per capita expenditure. Let W_j be a measure of the attractiveness of retail provision in zone j and c_{ij}, a measure of the travel impedance between i and j. The flows are then contained in the matrix $[T_{ij}]$ (where we are using T, rather than the more conventional S, to reserve S for other uses later). Then,

$$T_{ij} = A_i e_i P_i W_j^{\alpha} e^{-\beta c_{ij}}, \tag{1}$$

with parameters α and β representing the strength of attraction and of impedance, respectively.

$$A_i = \frac{1}{\sum_{j=1}^{N} W_j^{\alpha} e^{-\beta c_{ij}}}, \tag{2}$$

to ensure that the constraint

$$\sum_{j=1}^{N} T_{ij} = e_i P_i, \tag{3}$$

is satisfied. Since this is a singly constrained model, we can also calculate

$$D_j = \sum_{i=1}^{N} T_{ij} = \frac{\sum_{i=1}^{N} e_i P_i W_j^{\alpha} e^{-\beta c_{ij}}}{\sum_{k=1}^{N} W_k^{\alpha} e^{-\beta c_{ik}}}, \tag{4}$$

as total revenues attracted to retail in a zone, substituting for A_i in T_{ij} from (2)[2].

The set of j-zones, the retail centres, is likely to be different from the set of i-zones, the residential areas housing the consumers[3].

3 | Reducing Dimensionality: The Subsets

As a key step to reducing dimensionality, we set up a two-tier spatial system. A simple case would be a single city region, embedded in a system of super-regions within a country—for example, the United Kingdom. As we noted in Section 1, this is likely to necessitate a very high-dimensional set of variables. To reduce the dimensionality, we focus on significant flows. To achieve this, we first specify subsets of principal flows which help us to meet the dimensionality challenge. For this simple case, the first step is to reduce the number of super-regions to a subset that capture only the significant flows in or out of the region of interest. The criterion of "significance" can simply be a minimum size of flows, the T_{ij}'s in Equation (1) above, which might be determined experimentally. This is a principle that can be refined through further research.

To proceed, we develop a notation for a more general system with several regions of interest, each of which can have a distinct set of super-regions. For example, the set of regions could be the set of "cities" within the United Kingdom, each of which would have a set of super-regions defining their hinterlands.

As the basis of a general notation, we define R as a set of regions of interest, with R^r as the r^{th} region within the set. Let S be a set of super-regions, S^{Rr} be the subset that represent the hinterland of R^r, so that S^{RrSs} is the s^{th} such super-region.

Zonal variables, such as population would be, for example, P_i^{Rr} or P_j^{RrSs}—the population in the i^{th} zone of R^r in the first case, and in the j^{th} zone of the S^{RrSs} super-region associated with R^r in the second.

In specifying interactions either within the region of interest, or between the region of interest and its associated super-regions, we divide the two superscripts by a comma. The possible types of interactions are as follows:

within R^r, with Δ as the minimum size of a flow

$$T_{ij}^{Rr,Rr}, \quad i \in R^r, j \in R^r, T_{ij}^{Rr,Rr} > \Delta, \tag{5}$$

and for R^r to S^{RrSs}

$$T_{ij}^{Rr,RrSs}, \quad i \in R^r, j \in S^{\text{RrSs}}, T_{ij}^{Rr,RrSs} > \Delta, \tag{6}$$

and for S^{RrSs} to R^r

$$T_{ij}^{RrSs,Rr}, \quad i \in S^{\text{RrSs}}, j \in \text{R}^{\text{r}}, T_{ij}^{RrSs,Rr} > \Delta. \tag{7}$$

If necessary, we can also add flows within S^{Rr} and between S^{RrSs} and $S^{\text{RrSs}'}$, though these will probably be neglected in most applications. For generality, we do show how these can be included in Section 4.2 illustration.

The various flow possibilities are shown in Figure 1. The internal zonal structures of either regions or super-regions can be designed as appropriate for each case: A square grid for regions and super-regions can have a finer-scale such grid for the internal structure; if the regions are geographies, for example, Census areas or local authorities, the internal structures could be LSOAs in the Census case or local authority wards in the second case. In Figures 2, 3, 4, and 5, we show Case 1 (lower-level zones and super-regions with a predefined hierarchical relationship), and Case 2 (where the region of interest is defined relative to the selected i-zone), for a grid system (A) or a system of geographies (B).

In this simple case (Figure 2), one of the super regions is the single region of interest, R^r and with a square grid system of zones. The left-hand side shows the selection of an i-zone within R^r; all super-regions are shown with a zone structure. On the right-hand side, the super-regions connected to R^r and shown with the internal structure collapsed to a zone centroid.

Case 1B (Figure 3) is the same in principle as 1A but with the zone system as real geographies rather than grid squares. LSOAs are used in this illustration with R^r as an area in Manchester and a selected i-zone within it [9].

Cases 2A and 2B are essentially identical to 1A and 1B but adapt the boundaries of an existing set of super-regions post i-zone selection. We can run through a number of regions, R^r, where r = 1, 2, ... , N for each R^r, and the relevant super-regions are the sets S^{Rr} defined earlier.

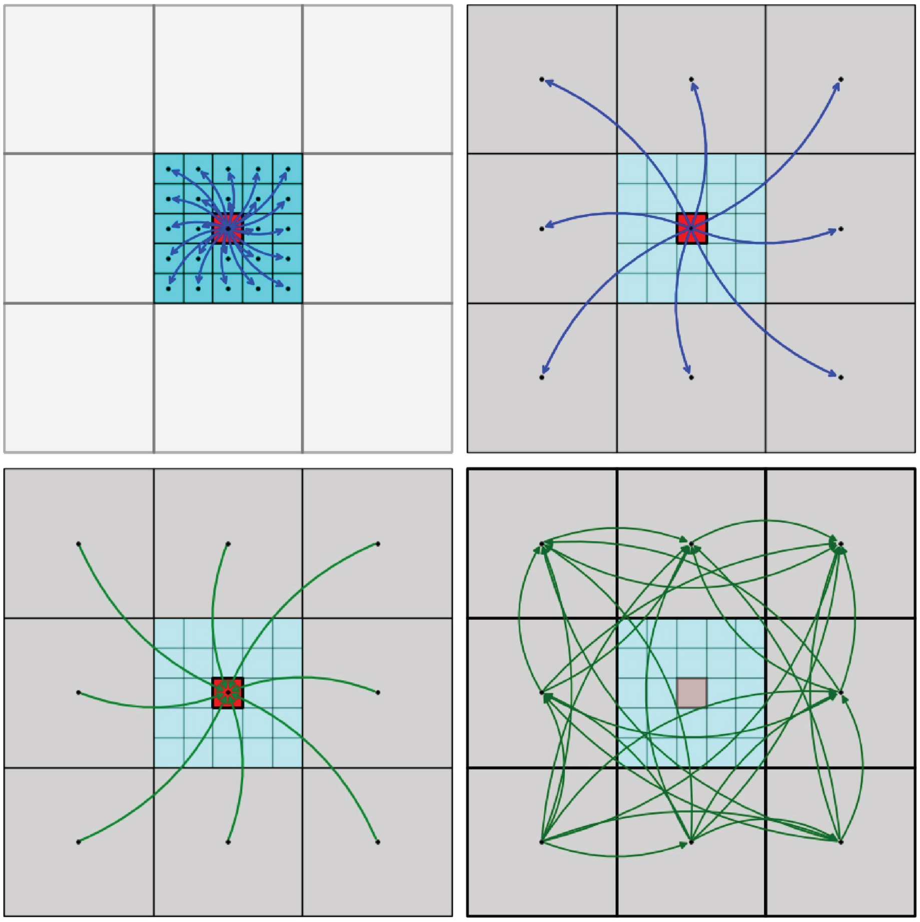

FIGURE 1 | The four potential types of flows in the system: (top-left) flows within R^r; (top-right) flows from R^r to S^{Rr}; (bottom-left) flows from S^{Rr} to R^r; (bottom-right) flows between S^{Rr} super-regions.

Case 2A is shown in Figure 4. The left-hand side shows a definition of R^r for a square zone grid system from an i-zone within it. The right-hand side shows how this overspills into nearby super-zones which then shrink in size. The S^{Rr} (modified) super-zones show with zone centroids (though they could also have a more complex internal zone structure, if necessary).

Figure 5 shows Case 2B. As above, this is similar to Case 2A but with real-geography zones—a version of the same LSOA within Manchester but with R^r defined in a different way (in this case, a distance buffer). The important point is that the method is sufficiently flexible to utilise a set of existing spatial hierarchies, or define/redefine these dynamically.

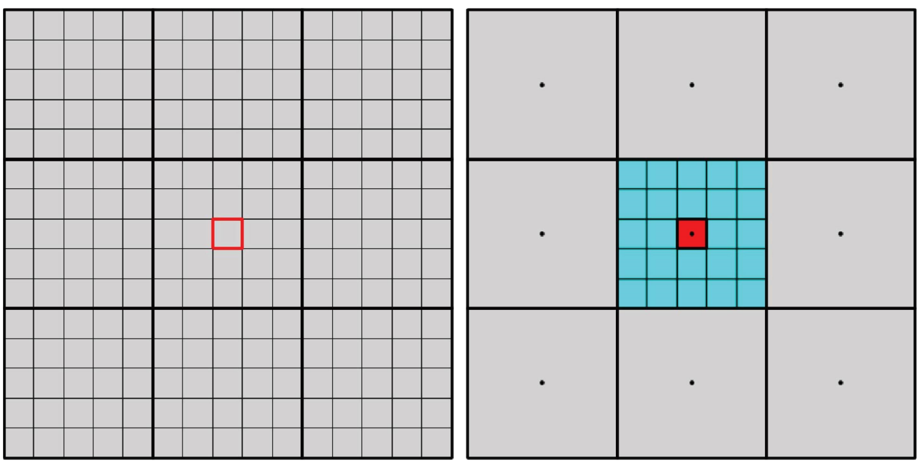

FIGURE 2 | An example of Case 1A for a square grid system: (left) the original system consisting of regions and super-regions, with the zone of interest given in red; (right) the reduced system, with the zone of interest within R^r, and simplified S^{Rr} regions.

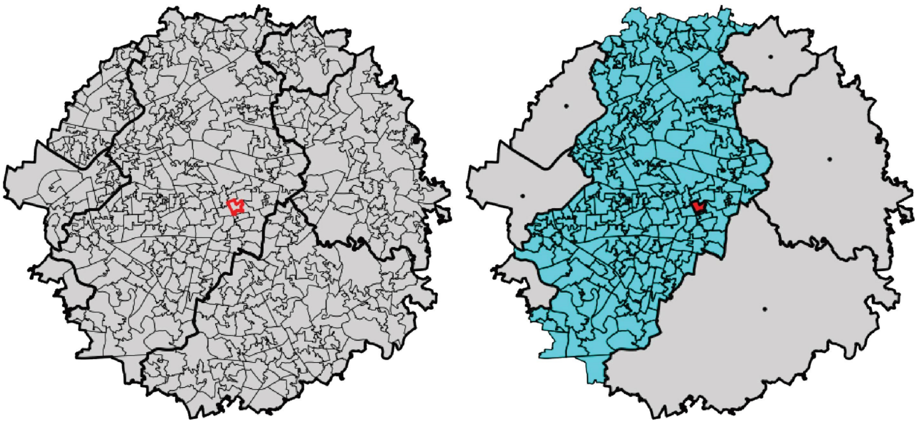

FIGURE 3 | An example of Case 1B with real geographies: (left) the original system consisting of regions and super-regions, with the zone of interest given in red; (right) the reduced system, with the zone of interest within R^r, and simplified S^{Rr} regions. Boundary data from ONS [9].

4 | The Equations for the Retail Model

4.1 | The Simple System

We first take a simple case, elaborating the model presented in Section 2. We re-present Equations (1)–(4), in terms of subsets to restrict the number of modelled flows, to demonstrate how we show only the principal flows. We also note that there is a further reduction in dimensionality by only including flows that exceed a certain minimum size, Δ (this is an assumption, as we noted earlier, that will need further refinement). Recall that R is the set of regions of interest and R^r is the r^{th} such region. We are assuming that each R^r is connected to a limited number of super-regions, the subset, S^{Rr}, and that each super-region subset has a single i- or j-zone, the centroid. Each R^r is assumed to interact with a subset of S^{Rr}'s, but the S^{RrSs}'s do not interact with each

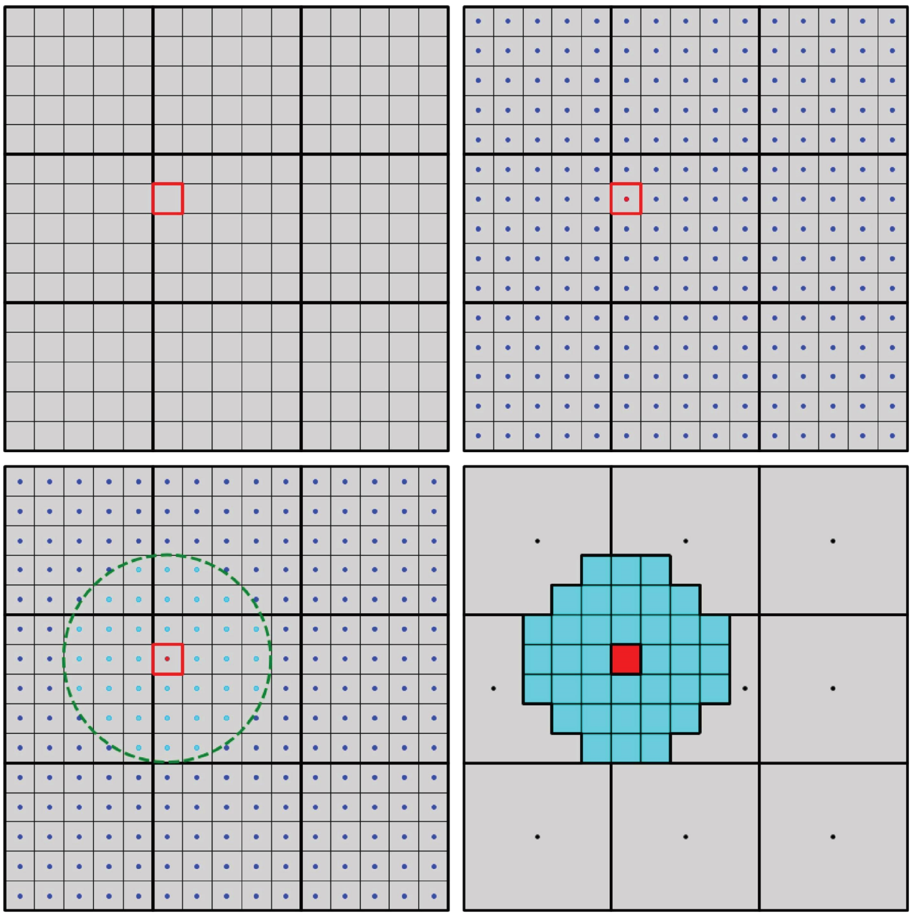

FIGURE 4 | An example of Case 2A: (top-left) the selected i-zone; (top-right) the centroids of all lower-level zones; (bottom-left) the R^r super-region, defined relative to the selected i-zone; (bottom-right) the adjusted set of S^{Rr} super-regions.

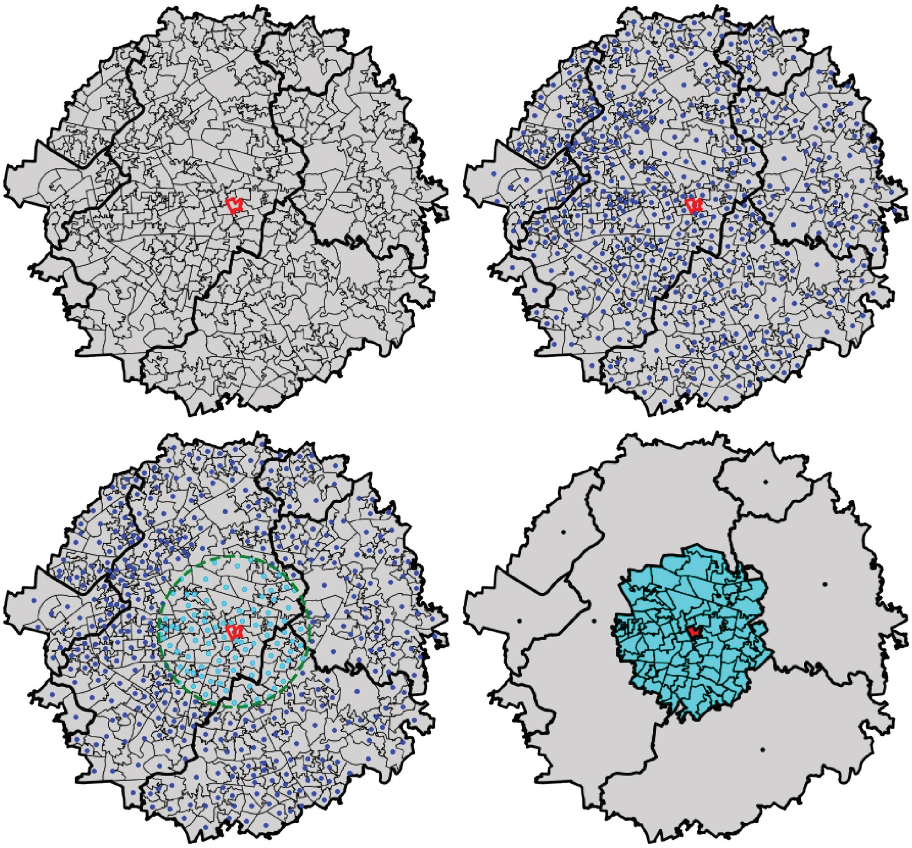

FIGURE 5 | An example of Case 2B: (top-left) the selected *i*-zone; (top-right) the centroids of all lower-level zones; (bottom-left) the R^r super-region, defined by a distance buffer relative to the selected *i*-zone; (bottom-right) the adjusted set of S^{Rr} super-regions. Boundary data from ONS [9].

other. The model can then be articulated as follows, beginning with flows internal to R^r:

$$T_{ij}^{Rr,Rr} = A_i^{Rr} e_i^{Rr} P_i^{Rr} W_j^{Rr\alpha} e^{-\beta^{Rr,Rr} c_{ij}^{Rr,Rr}}, \quad i \in R^r, j \in R^r, r \in R, \quad T_{ij}^{Rr,Rr} > \Delta. \tag{8}$$

Second, the flows from the *i*-zones in R^r to the *j*-zones in S^{RrSs}

$$T_{ij}^{Rr,RrSs} = A_i^{Rr} e_i^{Rr} P_i^{Rr} W_j^{RrSs\alpha} e^{-\beta^{Rr,RrSs} c_{ij}^{Rr,RrSs}}, \quad i \in R^r, j \in S^{Rr}, r \in R^r, \quad T_{ij}^{Rr,RrSs} > \Delta, \tag{9}$$

where A_i^{Rr} has to be calculated to ensure that all the $e_i^{Rr} P_i^{Rr}$ is spent, that is

$$\sum_{j \in Rr} T_{ij}^{Rr,Rr} + \sum_{j \in \mathrm{RrSs}} T_{ij}^{Rr,\mathrm{RrSs}} = e_i^{Rr} P_i^{Rr}, \quad i \in R^r, s \in S^{\mathrm{RrSs}}. \tag{10}$$

So that

$$A_i^{Rr} = \frac{1}{\sum_{j \in Rr} W_j^{Rr\alpha} e^{-\beta^{Rr,Rr} c_{ij}^{Rr,Rr}} + \sum_{j \in \mathrm{RrSs}} W_j^{RrSs\alpha} e^{-\beta^{Rr,RrSs} c_{ij}^{Rr,RrSs}}}, \quad i \in R^r, r \in R^r. \tag{11}$$

We then specify the flows from *i*-zones in S^{RrSs} into *j*-zone in R^r s

$$T_{ij}^{\mathrm{RrSs},Rr} = A_i^{\mathrm{RrSs}} e_i^{\mathrm{RrSs}} P_i^{\mathrm{RrSs}} W_j^{\mathrm{Rr}\alpha} e^{-\beta^{\mathrm{RrSs},Rr} c_{ij}^{RrSs,Rr}}, \quad i \in S^{\mathrm{RrSs}}, j \in R^r, s \in S^{\mathrm{RrSs}}, r \in R, \quad T_{ij}^{RrSs,Rr} > \Delta. \tag{12}$$

The A_i^{RrSs} s are a special case here as we are not estimating other flows from $i \in S^{\mathrm{RrSs}}$. Let e_i^{RrSs} and P_i^{RrSs} be the expenditure and population in S^{RrSs} purchasing in R^r. Then, we have

$$e_i^{\mathrm{RrSs}} P_i^{\mathrm{RrSs}} = \sum_{j \in Rr} T_{ij}^{\mathrm{RrSs},Rr} = A_i^{\mathrm{RrSs}} P_i^{\mathrm{RrSs}} \sum_{j \in Rr} W_j^{\mathrm{Rr}\alpha} e^{-\beta^{\mathrm{RrSs},Rr} c_{ij}^{\mathrm{RrSs},Rr}}, \tag{13}$$

so that

$$A_i^{\mathrm{RrSs}} = \frac{1}{\sum_{j \in Rr} W_j^{\mathrm{Rr}\alpha} e^{-\beta^{\mathrm{RrSs},Rr} c_{ij}^{\mathrm{RrSs},Rr}}}. \tag{14}$$

The total flows into a *j*-zone in R^r are

$$D_j^{Rr} = \sum_{i \in Rr} T_{ij}^{Rr,Rr} + \sum_{i \in Rr} T_{ij}^{Rr,\mathrm{RrSs}} + \sum_{i \in \mathrm{RrSs}} T_{ij}^{\mathrm{RrSs},Rr}. \tag{15}$$

If $[c_{ij}]$, $[e_i]$, $[P_i]$, and $[W_j]$ in their various forms are specified exogenously, then the model can be run for a set of R^r regions one

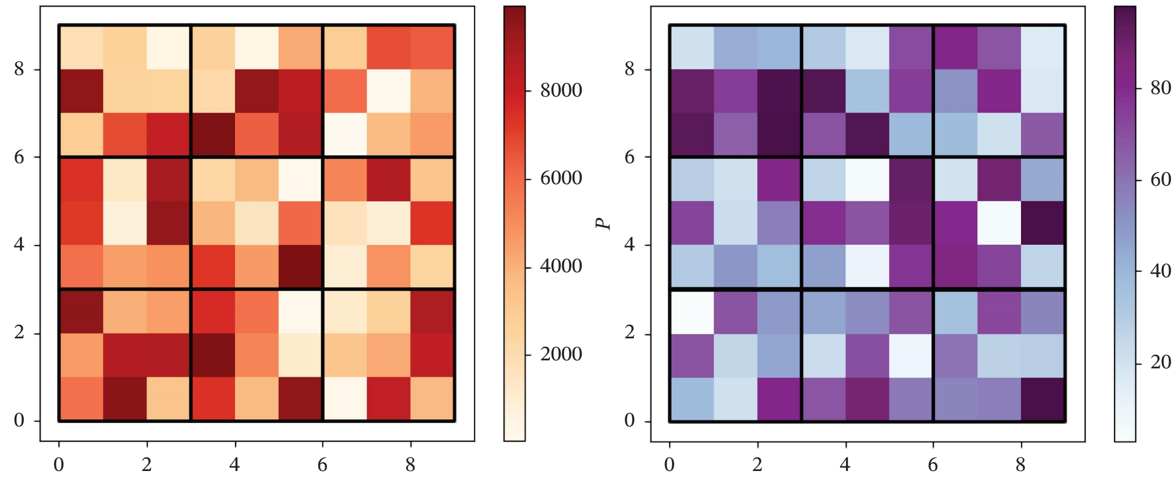

FIGURE 6 | The spatial distribution of values for P (left) and W (right). Each small square represents a lower-level unit (i), and each 3 by 3 square denotes a super-region (S^{RrSs}).

region at a time. If, however, $[W_j]$ is determined by the relation representing the equilibrium of a dynamic model

$$W_j^{Rr} = KD_j^{Rr}, \tag{16}$$

for appropriate W_j s, and for some constant, K, then the model would be run iteratively over $R^r \in R$. At the end of a cycle, all the W_j s, in each of R^r and S^{RrSs} would be adjusted through Equation (13).

4.2 | The General System

We can now in principle progress to a more general case. W_j, $j \in S^{\mathrm{RrSs}}$ may be dependent on other j-zones within S^{RrSs} and j-zones in S^{RrSs}'s other that those specified in S^{Rr}. We now allow the full generalisation by allowing for all the interactions in each S^{Rr}, and between S^{RrSs}'s (so that S^{RrSs} has a full set of j-zones, not just the centroid).

We can extend the notation by including $T_{ij}^{\mathrm{RrSs,RrSs}'}$ to be the set of i-j pairs within S^{RrSs} which are significant for i-j flows in S^{RrSs}. The full specification of this extension, which will not usually be needed, is left as an exercise, building on Equations (8)–(16) in Section 4.1.

5 | The Numbers: Dimensionality Reduction

The goal is ultimately reduce the number of calculations necessary to model the system. Suppose we have 10 super-regions, S^{RrSs}, and 40 regions, R^r. Regions and super-regions each have 100 zones. There would be $(40 \times 100) + (10 \times 100) = 4000 + 1000 = 5000$ zones in all. To model, all interactions would be $5000 \times 5000 = 25,000,000$, or 25 million. If the system is run for 10 iterations, this becomes 250 million (and often it would be necessary to run for more iterations than this). In this section, we will show how it is possible to reduce system complexity through our subset definitions—we address the preservation of information despite this simplification in the next section.

Consider the following:

A single region with 100 i-zones, and 100 j-zones

Each i-zone interacts with all 10 super-regions through their centroid (or some other representative point)

The number of principal flows in the system is therefore:

- T_{ij} within $R^r = 100 \times 100 = 10000$
- T_{ij} between R^r and $S^{Rr} = 100 \times 10 = 1000$
- T_{ij} between S^{Rr} and $R^r = 10 \times 100 = 1000$

So, 12,000 interactions for a computational run of a single region; or for the whole system with 40 regions, this becomes 480,000 interactions. If the system is run iteratively, for 10 iterations, there would be 4.8 million computations—which is significantly less than the original 250 million. This number might be increased if interactions between super-regions are considered, but these are relatively small due to their coarse nature.

6 | A Worked Example

To illustrate the viability of this approach, we provide a worked example. For our simplified test case, we create a set of random flows in a 9 by 9 grid, with square cells of equal size (see Figure 6). Each cell belongs to a super-region (S^{RrSs}), which in this case are larger square blocks of 9 cells. Each cell is randomly allocated a value for P between 1 and 10,000, and W between 1 and 100. For the sake of simplicity, we set e equal to 1—although this can be relatively easily incorporated, so long as the values are scaled to avoid violating Equation (3).

For the "full" model, any lower-level cell may interact with all other cells, giving us 81 by 80 combinations (excluding instances where $i = j$), or 6480 flows in total (Figure 7). In practice, this is small enough that calculating all flows is computationally trivial but is sufficient to demonstrate how our method works.

We then apply our method of reducing the model complexity, by grouping all cells outside of R^r into their respective S^{RrSs}. As is evident from Figure 8, this creates a much simpler flow pattern.

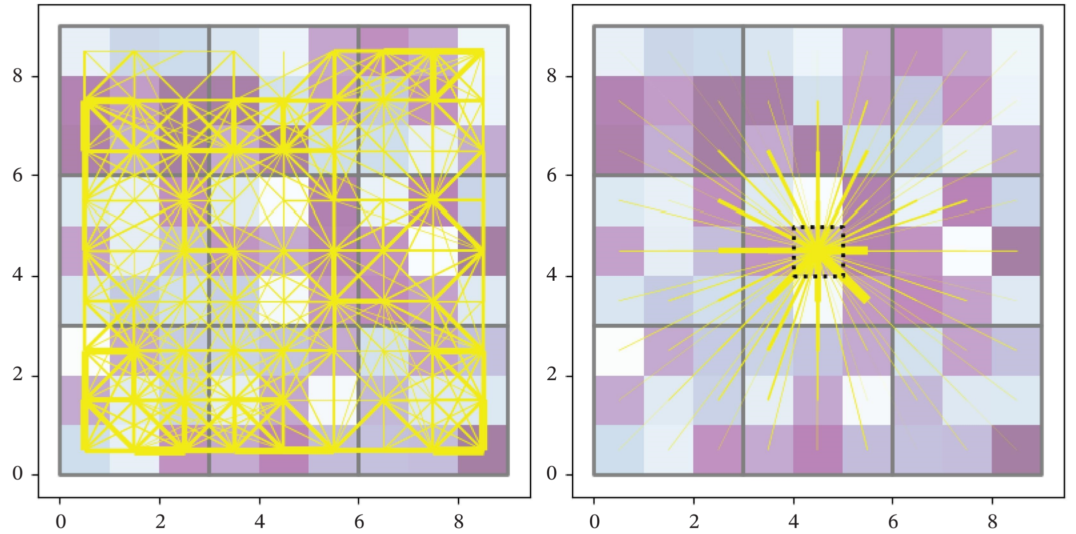

FIGURE 7 | The full model. Super-regions are indicated in grey for reference but have no effect on flows in the full model. Note: line width is scaled proportionally to the maximum flow observed in each diagram, and therefore is not directly comparable between subplots. (left) Flows in the full model, overlayed over values for W. For readability, only the top 10% of flows are shown. Line width is proportional to flow size; (right) flows in and out of the zone of interest in the full model, overlayed over values for W. All flows are shown. Line width is proportional to flow size.

To compare the full model to our reduced version, we utilise standardised root mean squared error (SRMSE), mean absolute percentage error (MAPE) and R-squared (coefficient of determination) as metrics. Each provides a different perspective on error in the reduced model, relative to the full set of flows—which we consider at the super-region level, for consistency in spatial scale. We opt for scale invariant metrics in order to facilitate comparability between iterations with different P and W values (discussed below).

These metrics can be defined as

$$\mathrm{SRMSE} = \frac{\sqrt{(1/n)\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\bar{y}}. \tag{17}$$

Values of SRMSE closer to 0 indicate that a model is a good fit. There is no theoretical upper bound, but higher SRMSE values indicate the average error between predictions and the underlying flows is larger.

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \tag{18}$$

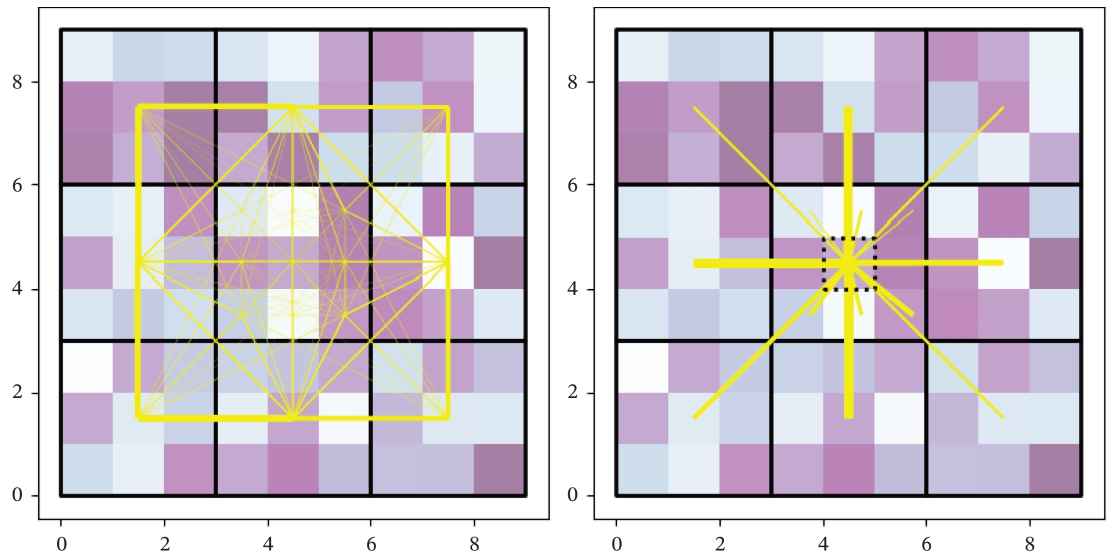

FIGURE 8 | The reduced model, with a single flow to/from each super-region. Note: line width is scaled proportionally to the maximum flow observed in each diagram, and therefore is not directly comparable between subplots. (left) Flows in the reduced model, overlayed over values for W. Line width is proportional to flow size. All flows are shown. Line width is proportional to flow size; (right) flows in/out of the zone of interest in the reduced model, overlayed over values for W. All flows are shown. Line width is proportional to flow size. Flows within R^r are offset for readability.

TABLE 1 | SRMSE, MAPE and R^2 comparing the fit of the reduced model to the full model over 100 iterations.

	SRMSE	MAPE (%)	*R*-squared
Mean	0.151554	8.9947	0.987739
Minimum	0.106504	6.4173	0.973270
25%	0.132989	8.0642	0.985240
50%	0.150544	8.9088	0.988389
75%	0.167509	9.8242	0.990427
Maximum	0.217819	12.1282	0.994889

Note: Metrics as implemented by scikit-learn [10].

MAPE gives the average absolute error (MAE) as a percentage of the observed flow. MAPE is typically presented as a percentage (0–100%). Lower values indicate a better model fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}. \tag{19}$$

The R-squared indicates the proportion of variance in the observed flows explained by the reduced model. *R*-squared is often considered as being bounded between 0 and 1, although it can take a negative value if model fit is particularly bad. Values closer to 1 demonstrate a good fit.

We then run the process for 100 iterations—with random values for P and W (provided in the supporting data file). For each iteration, we fit five parameters (α and β coefficients allowed to vary by level of the model: R^r to R^r α, R^r to S^{RrSs} α, S^{RrSs} to $S^{RrSs'}$ α, R^r to R^r β and R^r and S^{RrSs} to R^r β) using differential evolution to find the optimal values (defined as minimising the SRMSE), with each parameter bounded between 0.01 and 3.00. This allows us to approximate the error across different starting states for the model. The results are given in Table 1: MAPE and R-squared both indicate a good model fit, whilst the SRMSE values are (relatively) higher, they are still within an acceptable distance of the full model to be considered useful in high-dimensional settings.

7 | Next Steps and Further Work

As noted in Section 1, it would be possible to test the aggregation effects of the model system proposed here using alternative measures. It is also possible to extend the range of application—for example, beyond the retail model to the more general model rooted in input–output accounts [11]. In this case, the core model array is four-dimensional—x_{ij}^{mn}—for economic sectors (m, n) and a spatial system (i, j), and this makes dimensionality reduction even more important. This will be presented in a subsequent paper. It should also be remarked that the two-tier system developed here, and the related definition of sets of "significant" flows, could be extended to a n-tier, multilevel system[4].

Funding

No funding was received for this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available in the supporting information of this article.

Endnotes

[1]For this simple introductory illustration, we assume the same number of origin zones (i) and destination zones (j) and hence the summation are from 1 to N in each case. Generally, the numbers of i-zones and j-zones will differ and, with obvious definitions, the i-summations will be from 1 to N_1 say, and the j-summations, from 1 to N_2.

[2]To avoid confusion, since j is now not a free index on the right-hand side of Equation (4), we sum over a free index k in the internal summation on the right.

[3]For a model of retail flows in London, there may be around 600 (residential) i-zones and 200 j-zones (retail centres).

[4]The superscripts in this presentation are set labels that define the $\{i\}$, $\{j\}$ sets in the subscripts. In the later paper, we will move these to the subscript line and use the superscript line for the m, n economic categories.

References

1. T. A. Broadbent, "A Hierarchical Interaction-Allocation Model for a Two-Level Spatial System," *Regional Studies* 5, no. 1 (1971): 23–27, https://doi.org/10.1080/09595237100118503l.

2. G. Arbia and F. Petrarca, "Effect of Scale in Spatial Interaction Models," *Journal of Geographical Systems* 15, no. 3 (2013): 249–264, https://doi.org/10.1007/s10109-013-0180-9.

3. M. Batty and P. K. Sikdar, "Spatial Aggregation in Gravity Models. 1. An Information-Theoretic Framework," *Environment and Planning A* 14, no. 3 (1982): 377–405.

4. A. Hagen-Zanker and Y. Jin, "A New Method of Adaptive Zoning for Spatial Interaction Models," *Geographical Analysis* 44, no. 4 (2012): 281–301, https://doi.org/10.1111/1538-4632.2012.00856.x.

5. F. Roumpani, *The Role of Procedural Cities in the Future of Planning: An Integrated Method* (University College London, 2019), PhD thesis.

6. I. S. Lowry, *A Model of Metropolis* (The Rand Corporation, 1964), 1–34.

7. J. Dearden and A. Wilson, in *Explorations in Urban and Regional Dynamics* (Routledge, 2015).

8. A. G. Wilson, "A Family of Spatial Interaction Models, and Associated Developments," *Environment and Planning A* 3, no. 1 (1971): 1–32, https://doi.org/10.1068/a030001.

9. ONS Geography Office for National Statistics, "Lower Layer Super Output Areas (December 2021) Boundaries EW BFC (V10)," (2024), https://geoportal.statistics.gov.uk/datasets/ons::lower-layer-super-output-areas-december-2021-boundaries-ew-bfc-v10-2/about.

10. F. Pedregosa, G. Varoquaux, and A. Gramfort, "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research* 12 (2011): 2825–2830.

11. G. Solomon and A. Wilson, "The Mesolevel Economy in Nineteenth-Century England and Wales: Applying Input-Output Accounting and Spatial Interaction Modelling to the Historical Study," *Complexity* 2024 (2024): 3016105, https://doi.org/10.1155/2024/3016105.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Supporting Information.** Reducingdimensionality_Complexity_supdata.csv contains the supporting data for replication of each of the iterations discussed in Section 6.

Annex: Definitions of Symbols Used

In the illustrative retail model in Section 2.

N, number of zones in a spatial system

T_{ij}, the flow between the i^{th} residential zone to the j^{th} retail zone

e_i, expenditure per capital of residents in zone i

P_i, the population of zone i

W_j, a measure of the attractiveness of retail in zone j

c_{ij}, a measure of the travel impedance from zone i to zone j

α, a parameter which measures the strength of the attractiveness of a retail zone

β, a parameter which measures the strength of the travel impedance

A_i, a balancing factor to ensure that the total expenditure is allocated between retail zones

Sets: regions, super-regions and subsets.

R, a set of regions of interest

R^r, the r^{th} such region within R

S, a set of super-regions

S^{Rr}, the subset of super-regions interacting with the r^{th} region of interest in R^r

S^{RrSs}, the s^{th} subregion within S^{Rr}

Variable definitions for the disaggregated retail model (note: the relevant zone totals in the summations are determined by set inclusion signs, $\in$, with the subset definitions above).

e_i^{Rr}, per capita expenditure in the i^{th} zone of R^r

e_i^{RrSs}, per capita expenditure in the i^{th} zone of the S^{RrSs} super-region associated with R^r

P_i^{Rr}, the population of the i^{th} zone of the R^r region of interest

$P_i^{Rr,\text{RrSs}}$, the population of the i^{th} zone of the S^{RrSs} super-region associated with R^r

W_j^{Rr}, the attractiveness of the retail in the j^{th} zone of R^r

W_j^{RrSs}, the attractiveness of the j^{th} zone of the S^{RrSs} super-region associated with R^r

$T_{ij}^{Rr,Rr}$, the flow between the i^{th} residential zone of R^r to the j^{th} retail zone of R^r

$T_{ij}^{Rr,\text{RrSs}}$, the flow between the i^{th} residential zone of R^r to the j^{th} retail zone of the S^{RrSs} super-region associated with R^r

Δ, the minimum size of a flow to be used in the model

$c_{ij}^{Rr,Rr}$, a measure of the travel impedance from zone i in R^r to zone j in R^r

$c_{ij}^{Rr,\text{RrSs}}$, a measure of the travel impedance from zone i in R^r to zone j in the S^{RrSs} super-region associated with R^r

A_i^{Rr}, a balancing factor for flows from zone i in R^r

A_i^{RrSs}, a balancing factor for flows from zone i in the S^{RrSs} super-region associated with R^r

$\beta^{Rr,Rr}$, the travel impedance parameter for flows within R^r

$\beta^{Rr,\text{RrSs}}$, the travel impedance parameter for flows between R^r and the S^{RrSs} super-region associated with R^r