Translating analytical descriptions of cities into synthetic urban design models

Kinda Al Sayed
Alan Penn
University College London, UK

With the increase in urban complexity, plausible analytical and synthetic models became highly valued as the way to decode and reconstruct the organization that makes urban systems. What they lacked is a mechanism by which an analytical description of urban complexity could be translated into a synthetic description. An attempt to define such a mechanism is presented in this paper, where knowledge is retrieved from the natural organization that cities settle into, and devised in a design model to support urban design at the problem definition stage. The model comprises two automated design modules, giving preference to street accessibility. The first module implements plausible spatial laws to generate street structures. The performance criteria of these structures are measured against accessibility scores and clustering patterns of street segments. In the second module, an Artificial Neural Networks model (ANNs) is trained on Barcelona's data, outlining how street width, building height, block density and retail land use might be dependent on street accessibility. The ANNs is tested on Manhattan's data. The application of the two computational modules is explored at the problem definition stage of a design process in order to verify how far deterministic knowledge-based models are in the transition from the analysis of design problems to the synthesis of design solutions. Our findings suggest that the computational framework proposed could be instrumental at generating simplified representation of an urban grid, whilst being effective at forecasting form-related and functional attributes within a minimum resolution of 200 meters. It is finally concluded that as design progresses, knowledge-based models may serve as to minimize uncertainty about complex urban design problems.

Design Computing and Cognition DCC'16. J.S. Gero (ed), pp. xx-yy. © Springer 2016

Introduction

Over the last decades, urban studies were witnessing a divide between the analytical sciences and the applied sciences of cities. Analytical sciences embraced many attempts to decode urban complexity by means of explanatory models [1], [2], [3], [4], [5], and [6]. A complementary effort was made in applied urban sciences, where more emphasis was laid on assumption-based simulation models on the scale of cities and regions [7], [8]. Any attempts to bridge the divide between the analytical and applied sciences of cities were faced by non-trivial challenges, perhaps for the very reason that backed critics against Alexander's work [9]; that is the inherent distinction between analysis and synthesis. To bridge between analysis and synthesis, there needed to be some intuition into the type of mechanism required to convert an explanatory description of urban phenomena into a synthetic design approach. How far can these explanatory descriptions be used in reconstructing urban complexity is a question that needs further investigation in the realm of design and computation. In response to this question, this paper embraces an attempt to encode a synthetic description of the organization that couples street structures with form-function attributes of urban fabric. Learning from Barcelona, Manhattan, and London, a knowledge-based model is devised to aid urban design.

In line with observed self-referential processes in street networks [10], and the premise that form-function attributes are dependent on the spatial accessibility of road infrastructure [2], [11], the model proposed here outlines a prioritized structure of design thinking, comprising two automated modules to enable the generation and evaluation of street networks, and the prediction of form-function attributes of urban structures. The methodological framework for the two design modules is explained in detail. The generative module is to utilize plausible rules and empirically-validated benchmarks for assessing the urbanity of the generated street structures. The forecasting module is to devise an artificial Neural Network model to forecast form-function features of the generated street structure. Both modules are then applied to generate simplified descriptions of an urban grid and predict its attributes. The application of these two modules serves as to explore the extent to which knowledge-based models might determine some features of an urban grid.

From analytical descriptions to design prescriptions

Bridging the gap between complexity and design, significant contributions were made within the framework of urban modelling and simulations [12] and [13]. For the most part, research in these domains came short of high resolution structural descriptions of urban form. To adapt such descriptions into the linear course of design, a comprehensive framework was required to decode, encode and reconstruct the architecture of cities. For the purpose of developing such a framework, there is a need to frame the problem definition of cities before tackling the problem of design.

One of the first calls to define urban problems was that of Jane Jacobs [14], where she called for understanding cities as problems of organized complexity. Any translations of this understanding into quantitative descriptions were subject to representation. In general, we could recognize two types of approaches; that of Space Syntax and that of complexity science. Space Syntax is a theoretical framework that builds on a hypothetical relationship between street structures, natural movement and socioeconomic processes [1], [2]. This theoretical proposition is debated in the context of complexity science. Complexity scientists and geographers often questioned the overreliance on linear models used to describe urban relationships [15] and [16], whilst questioning the validity of two dimensional representations of urban phenomena [17]. This is in view of the argument that reductionist models that relied on simple causal relationships between two variables or more were not immune to erroneous Similar skepticism was posed against urban simulation models [13], mainly questioning the over-reliance on assumptions in simulation models, especially when no clear explanation was given on how a plausible knowledge about cities could inform design synthesis. With some exceptions [18] and [19], the majority of computational urban design models were not directly based on tangible knowledge about the mechanisms that drive growth and differentiation in cities. Recently, there has been significant development on this front. Duarte et al. [20] have developed a computational urban design model called “City Induction”; which incorporated three sub-models; the first sub-model generated context-specific solutions based on the ontologies introduced in “Pattern Language” by Alexander et al. [21]. The second sub-model was based on Stiny’s shape grammar and description of designs [22], and was used to generate designs. The third sub-model was building on Hillier’s theory on space syntax [2] as an evaluative tool of urban design. There was no conceptualization, however, about how space syntax itself might be used to generate designs.

Attempts to inform urban design theory by virtue of empirical knowledge on spatiotemporal patterns of urban growth were more focused on the regional scale [23]. Methodologies varied depending on the computational models and the elementary seeds used in growth simulations. City blocks were often considered as the elementary components in these simulations [24] and [25]. Despite early attempts to combine structural and shape descriptions [26], there was generally less emphasis on simulating street structures. Some studies implemented L-Systems in procedural models utilizing discursive rules of addition and subdivision in streets [27], whilst other studies used accessibility scores to assess street patterns generated by means of agent-based models [28]. These studies, however, made no reference to empirical data on historical urban growth [10].

Forecasting models were also needed to cover a wide range of variables that represent urban complexity without looking at one or two variables in isolation of others. A modelling description of land use transformations in isolation of street infrastructure might pose serious challenges [29], particularly when measuring on a hypothetical dependency between street accessibility and urban form and function [2], [11], [30], and [31]. Similarly a separation between urban blocks and street patterns [32], might be questionable if we consider block agglomerations as the inverted representation of street spaces [33]. In the same way, a separation between street width and spatial accessibility [34] needs to be reconsidered when regarding street width as the supply for the demands of street network accessibility [11]. Urban design models might also benefit from a more comprehensive account of the relationships that couple transformations on blocks and changes on land uses, land values and building height as well as street spacing [35]. In reviewing research in this domain, the questions that persist are; how to simulate the growth of street structures in such a way as to build on temporal descriptions of urban transformations? And how to forecast form-function attributes of cities in such a way as to build on empirical models of urban structures and their dependencies?

A prioritised structure model for urban design

In response to the challenges presented in previous sections, this paper presents an attempt to build a synthetic description of space syntax at the problem definition stage of urban design, using a scheme that prioritizes urban structures. Our proposed model is based on two design filters. The first was a more defined version of the "generic function" [2]; that is a street network that is derived from local rules to generate a permeable street structure. The second filter was based on modelling the relationship between street accessibility and form-function variables (block density, street width, building height and land uses).

Methods for decoding and mapping spatial variables

In the next sections, we will briefly explain the methods developed in Space Syntax theory to measure spatial accessibility in streets. We will also describe our mapping methodology that enabled the construction of empirical models for the purpose of forecasting.

Network analysis of street spaces

In syntactic analysis, street structures are represented by topological and topo-geometric network representations, namely; axial maps and segment maps. An axial map is a network representation of the longest and fewest lines of sight that cover all street spaces. The segment map is a broken description of the axial representation where each segment element between two street inter-junctions is considered as a "node" in a street network, where intersections are considered to be links. In this network representation, nodes are spatially distributed and links are associated with a cost of turning from one street to the other [36]. Space Syntax research incorporates different measures of network distance; topological, metric and angular [37]. For the purpose of this paper, we were mostly concerned with angular depth and connectivity (degree) of street lines. Angular measures were proven to be powerful at capturing vehicular and pedestrian movement potentials as well as at highlighting catchment areas for active economic centers [38]. In segment analysis, integration (closeness) and choice (betweenness) measures can be used to capture the angular geometric properties of street networks. The angular network measures were not normalized until very recently [38], and are still under testing.

Mapping and aggregation techniques

In order to map street network measures against other continuous and ordinal variables, a special technique was developed for aggregating spatial data in a separate layer; called the pixelmapper [39]. Using this method, indices of street accessibility and form-function attributes of urban areas were aggregated within square polygons (Figure 1). Larger polygons or cells imply that relationships were captured within a lower resolution. Data was binned in two overlapping polygon grid layers. The second polygon grid layer was shifted diagonally so that the end point of a polygon in the second grid layer was placed on the center of a polygon in the first grid layer. The highest values of both layers were filtered in a third polygon layer with double the resolution of the previous two grid layers. We applied this method on Manhattan to capture the linear correlation coefficient between street accessibility (average NAIN per grid square, or total segment connectivity per grid square) and density of blocks, commercial land uses, high-rise development, and street width. We used different grid resolutions (200, 400, 600, 800, 1000, 1200, 1400, 1600 m). The analysis yields 1000 m as the ideal grid resolution for capturing the highest correlations between street accessibility and form-function variables. It was recognized, however, that lower resolutions will increase the risk of error and will misrepresent the local properties of urban structure [40]. For this reason, data was binned at both the (1000 m to 500 m), and (500 m to 250 m) resolution scales.

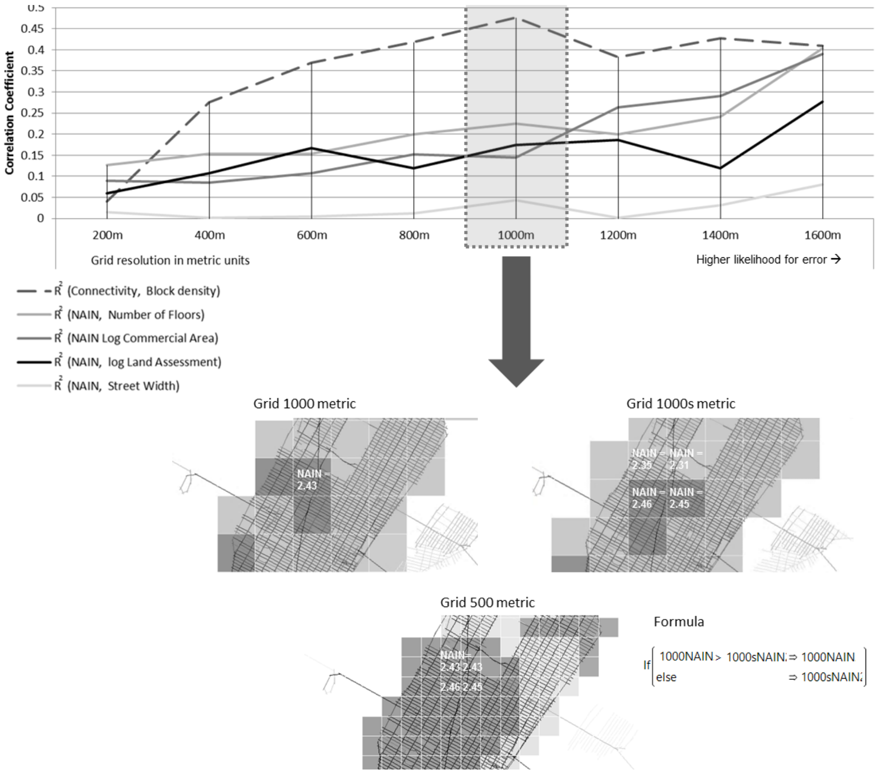


Fig. 1 Binning data to correlate street accessibility to street width and density of blocks, retail landuse and high-rise buildings. Different grid resolutions were used (200, 400, 600, 800, 1000, 1200, 1400, 1600 m), of which 1000 m was proving to score higher correlations. The method implies; storing data and spatial configurations in two overlapping grid reference layers and selecting the highest values in a third higher-resolution reference layer.

A modelling framework for semi-automated urban design

This paper will expand on the first two design filters through automating them in two separate modules; a generative one and a forecasting one. In the generative module, knowledge is utilized at two stages; during the implementation of the generative algorithm and at the evaluation stage. To simulate growth mechanisms, a set of spatial laws were applied to govern the length and angularity of street spaces. The generated outcome is then evaluated against certain spatial scores.

For the forecasting module, a nonparametric Artificial Neural Networks (ANNs) model is devised. The model relies on empirical data that define the relationship between street structures and form-related and functional attributes, including; street width, building height, block density and land uses in Barcelona and Manhattan.

A generative design module

Early Space Syntax experiments [1] presented a generative pattern of organization on the local scale of an urban area. The approach was further developed by Coates [41]. As to reflect on the emergent nature of the resultant grid structures and block alignments, Hillier [2] recognized the tendency of longer lines to continue straight and shorter lines to be blocked forming near-right angles. By identifying that process as the "centrality and extension" rule he made the assumption that global patterns of urban structures are an emergent product of local rules. Whether a centrality and extension rule on its own can lead to the generation of city structures is something that needs to be questioned, provided evidence on different feedback mechanisms that govern urban growth behavior [10]. At this stage, it is difficult to rule out the sequence in which these laws generate urban structures. We therefore take their overall features as criteria for urban pattern recognition.

Rules for generating street structures

In the first design module, we generated a number of growth iterations for hypothetical urban structures using Hillier's "centrality and extension" rules whilst allowing for a margin of randomness. Longer lines were encouraged to continue and intersect with other lines forming semi-continuous patterns. Shorter lines were more likely to stop at the first line they intersected with forming near-right angles where possible. The pseudo code follows the following logic;

- Draw three lines starting from random points within the screen area and following random directions;
- Divide each of these lines to 20 segments, and choose randomly one of the points of division as a seed for a new line;
- For 2% of the cases, if the original line happened to be longer than 400, and the new generated line started close to one of the end points of the original line, direct the new line in an angle that is within (0, ±12.8) degrees range.
- For another 2% of the cases, let the new line go in a direction that is within (0, ±42.8) degrees range.
- If the original line is shorter than 400, for 52% of the cases let the new line go in a direction that is close to 90 degrees ±2.2

The resultant structures presented varying syntactic properties. In order to recognize structural patterns that match those of cities we compared these iterations to real urban structures.

Assessing the urbanity of the growth iterations

To establish benchmarks for evaluating generative growth iterations, we compared the generated structures to a random structure and to London's street structure and an existing sample taken from Barcelona's deformed grid. The random structure is regarded as hypothesis null; marking the lowest performance of a street structure. The hypothesis is that structures generated by virtue of Hillier's simple rules will be more similar to real cities than to a random structure.

Through conducting research on 50 US cities, research by Bin Jiang confirmed that connectivity (degree) of street networks follows a power-law distribution [42]. This was also observed in the historical growth of Barcelona, which revealed preferential attachment dynamics [1]. This observation is used here to evaluate the structures of the generated iterations. From Table 1, it is clear that the correlation coefficient of power-law distributions is not a strong discriminator of real street networks compared to random networks. Iterations 3 and 1 presented closer values of correlation coefficient to both London and `Barcelona, but these values were also close to a random network. When measuring on the parameters of power-law distributions; *a* and *k*, the distinctions between random and urban systems became more visible. The values of *a* yielded iteration 4 as the closest iteration to Barcelona's deformed grid, whilst iteration 3 came second. The values of *k* yielded iteration 1 as the closest iteration to Barcelona and London, and again iteration 3 came second.

As established in [43], in a grid that presents a differentiated structure integration values tend to follow lognormal distribution. The distribution differs from that of random networks in that it shows a higher degree of skewness (asymmetry). On aggregate, the closeness centrality of a random network will be characterized by a normal distribution with minimal skewness. This constitutes the second benchmark for evaluating the urbanity of generative structures. The lognormal distribution can be evaluated through measuring the goodness-of-fit D representing the distance between the cumulative distribution and a cumulative fraction plot for the data sample. The goodness-of-fit is measured by running the empirical distribution function KSL test [44]. Skewness is also added as an indicator to the degree of asymmetry in the structure as a whole in comparison with a random structure.

Measuring on the cumulative structural properties of depth in the network, aggregate integration values did not seem to fit very well to a lognormal distribution compared to the randomised map and Barcelona. Barcelona's structure prevailed over a randomised map in its fitness to a lognormal distribution, and it showed higher degree of skewness.

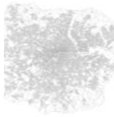
Judging on KSL test, iteration 3 presented a better fit with lognormal distribution as well as a differentiated structure (Skewness=-1.46).

Given that distributions do not interpret structural properties, another criterion is added to compare the relationship between axial connectivity and axial integration as a measure of intelligibility. Urban systems exhibit relatively high intelligibility between the local and global axial structures [45].

Considering intelligibility as a measure of the part-whole structural unity, the structure of iteration 3 was found to be more intelligible compared to other iterations. Yet, it is difficult to foresee how intelligibility might act as a law for recognising the urbanity of street structures, since our observations indicate that random networks are more intelligible than both Barcelona's grid, and London's street network. It is worth mentioning here that intelligibility is largely influenced by the system's size. To verify these results, we may need to use different rules and seeds for the randomised networks.

Considering these findings, iteration 3 prevailed as it presented an optimum foreground structure that conserved physical distance and angular turn costs. It also presented a higher level of structural differentiation that made a better match with real cities. On aggregate, angular depth values in iteration 3 followed a lognormal distribution. The structure of iteration 3 was also more intelligible than other structures. Despite the relative success of iteration 3, it failed to be fully compatible with real urban structures. Additionally, it was difficult to identify an optimum performance for the growth iterations, a performance that might fully comply with how urban structures are configured in real cities. However, for the purpose of our experiment; we proceeded by applying the ANNs model on iteration 3 to further define the design features of the generated structure.

Table 1 Evaluating the four growth iterations against the spatial properties of Barcelona and a randomly generated structure. The generative code is written in Processing (Java). Spatial Structures are analysed using UCL Depthmap [46].

		Iteration 1	Iteration 2	Iteration 3	Iteration 4	Barcelona	London (all)	Random structure
Spatial structure								
Power-law distribution of connectivity	R^2	0.92	0.93	0.92	0.926	0.865	0.832	0.87
	a	130.2	178.0	179.2	249.5	234.62	1251.7	1505.7
	k	-0.7	-0.72	-0.71	-0.73	-0.61	-0.579	-1
Lognormal Distribution	KSL	0.09	0.15	0.088	0.1	0.035		0.04
	Skewness	-1.32	-1.15	-1.46	-1.5	0.28		-0.18
Intelligibility (R^2)		0.12	0.13	0.17	0.1	0.33	0.061	0.56

A city-to-city learning approach

In this section, a supervised machine learning model will be applied using a soft computing technique based on ANNs. The use of ANNs in modeling would enable empirical encoding of data on space, form and function. The ANNs allow for minimizing assumptions about the input and output data distribution and the type of data used, whether continuous, categorical, or binary. They are particularly useful in cases where complexity in the system relationships and imprecision in observations are issues that threaten the credibility of simpler models. ANNs are also fault tolerant towards redundant information coding, where there are hidden relationships between spatial measures or between socioeconomic variables.

ANNs consist of layers and neurons that simulate human learning. The training of ANNs can help storing embedded functions that are then used to categorize information and provide projections given new situations. With such functionality, ANNs can be used to answer *what if* questions and generalize complex relationships on presumably similar situations to the situation used in the training. ANNs are used in many fields; including medical sciences, engineering, A.I. and many others. They are also known to be successful in the nonlinear mapping and modeling in geography and planning [47]. The downside in using ANNs is in the difficulty to describe the relationship between the input variables and the output variables. All the training takes place within a *black box*. Neural Networks comprise a large class of different model architectures. Traditionally, ANNs are used to classify a set of observations. In most cases, the issue is in approximating a static nonlinear, mapping $f(\mathrm{x})$ with a neural network $f(\mathrm{x})_{\mathrm{NN}}$, where $\mathrm{x} \in \mathrm{R}^{\mathrm{K}}$. The ANNs model to be used in training space and form-function data in this section will consist of three layers, the input, output and a layer with hidden nodes in-between. The different layers are encoded in the *multilayer-perceptron* (MLP) model[1] [48] illustrated in (Fig. 2). Three hidden nodes are considered in the middle layer, where activation functions that store weights and biases are embedded. The ANNs model will be *fully connected* and will use a *feed-forward* mechanism. The network is *fully connected* since the output from each input and hidden neuron is distributed to all of the neurons in the following layer. The *Feed forward* mechanism of the model entails that the values would only move in the forward direction from input to hidden to output layers; so that no values are fed backwards to input or hidden layers. Due to the limited number of inputs (3) and outputs (4) and a fair amount of redundancy (correlation) between two spatial measures in the input layer, we chose simple network architecture for the model.

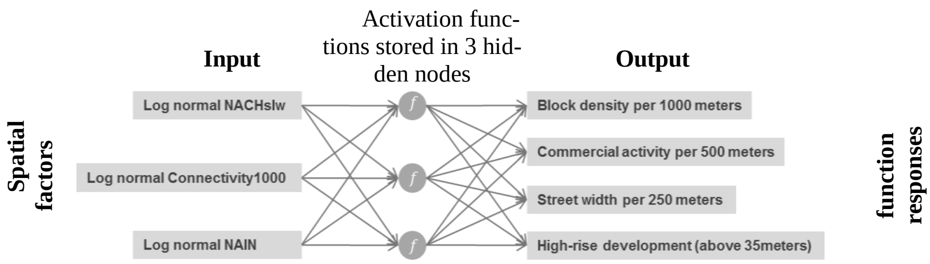


Fig. 2 An ANNs model applied to Barcelona and Manhattan, considering normalized spatial measures of choice, integration and connectivity as factors and form-function attributes as responses.

The ANNs is fitted using standard nonlinear Least-Squares Regression methods. The inputs x_n, $n = 1, \ldots, n$ to the neuron in the hidden layer are multiplied by weights w_{ni} and summed up together with the constant bias term Q_i. The resulting n_i is the input to the activation function y. The activation function used here is the hyperbolic tangent function which is a

1 MLP consists of multiple layers of simple, two state, sigmoid processing nodes/neurons that interact using weighted connections.

sigmoid function. It transforms values to be between -1 and 1, and is the centered and scaled version of the logistic function. The hyperbolic tangent function is:

$$f(\mathrm{x}) = \tanh(\mathrm{x}) = \frac{e^{2x}-1}{e^{2x}+1} \quad (1)$$

Where x is a linear combination of the X variables.

The output of node *i* is then defined as the following;

$$y_{\mathrm{i}} = f\left(\sum_{i=1}^{k} w_{ik}x_J + Q_i\right) \quad (2)$$

To explore the application of ANNs in urban design, we encoded empirical data from Barcelona, and tested the model against data from Manhattan. Geometric measures of street network configurations were used as inputs. The output was a combination of Form attributes (building height and density, street width), and functional (overall commercial zoning).

The ANNs were trained and validated against Barcelona's. In instructive training, the error information was propagated backwards through the network using a backward propagation algorithm. The algorithm iteratively minimizes an error function over the network outputs and desired outputs [49]; [50]. We used the KFold method to validate the operative mechanism of the model by recursively selecting one subset out of five. The subset that best validated the model was then chosen. The validation helped detecting if the model overfits the data.

For the input layer, we used normalised choice [Segment length weighted] (NACHslw), normalised integration (NAIN) and aggregate connectivity per 1000 square unit (Connectivity1000). All indices were computed using Depthmap [46]. NACHslw is an angular measure of graph betweenness that is normalised and weighted by street segment length [36]. NAIN is a normalised and angular-weighted measure of graph's closeness [38]. Connectivity is equivalent to degree in graph theory. It is here aggregated per 1000metre square unit. Both NACHslw and NAIN were calculated for the whole system (radius *n)*. Before using the continuous variables as input in the ANNs model, we normalised their values using a lognormal probability function to fit in the range [0, 1]. The dependent responses were a mix of continuous variables running in *regression* mode (Block density per 1000 metric square) and ordinal variables running in *machine* mode (commercial activity, street width above 30 meters, high rise above 35 meters). The positive presence of the ordinal response variables was marked as 1 and the negative presence is 0.

The performance of ANNs on Barcelona was evaluated using a Linear Regression for block density. Both *accuracy* and AUC measures were applied to evaluate the predictive power of the ANNs running in *machine* mode. For *accuracy*, we calculated the rate of classified scores against total scores from the confusion matrix[2]. The Receiver Operating Characteristic (ROC) curve plotted the true positive rate (sensitivity)[3] on the vertical axis and false positive rate (specificity)[4] on the horizontal axis. For ROC, we calculated the area under the ROC curves (AUC). We then observed the cross-validated estimates of *accuracy* and AUC. In addition, the Root Mean Square Error (RMSE) between vaildated and training data was examined to check for overfitting.

Training Artificial Neural Networks on Barcelona's data

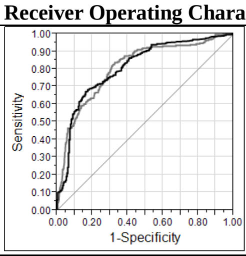
Measuring on *accuracy* and AUC, the ANNs were successfully fitted between the input (indices of accessibility) and the output (form-function) data. The difference in Root Mean Square Error (RMSE) between trained and validated data was minimal (0.01, 0.03, 0.01 for ordinal variables) showing no signs of overfitting. The AUC recorded values above 0.8, 0.81, 0.79 in predicting High-rise development, Commercial activity and Street width respectively (Figure 3). Measures of *accuracy* were recording 0.71, 0.82, 0.74 for classifying the presence of High-rise buildings, Commercial activity and Street width respectively. The correlation between actual and predicted block density was also high R^2=0.61. The evaluation rates indicated that spatial accessibility can classify the positive/negative presence of ordinal responses and correspond to block density.

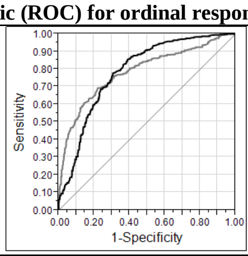
2 Accuracy can be calculated from the contingency table as follows;
((True Positives) + (True Negatives)) / ((True Positives) + (True Negatives) + (False Positives) + (False Negatives))

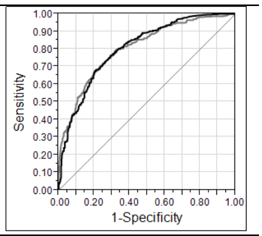
3 Sensitivity= True Positives / ((True Positives) + (False Negatives))

4 Specificity= False Positives / ((False Positives) + (True Negatives))

Receiver Operating Characteristic (ROC) for ordinal responses

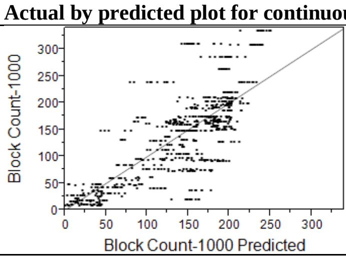






Commercial activity		**Street width**	**High-rise above 35m**
— **0** — **1**			
Area	0.81	0.79	0.80
Accuracy	0.82	0.74	0.71

Actual by predicted plot for continuous response



R^2 = 0.61

Block density per 1000metric square

Fig. 3 ROC and scattergram plots evaluating the performance of the Neural Networks on Barcelona's data. Spatial configurations were used as factors. Form-related and functional attributes were considered as responses

Testing the Artificial Neural Network model on Manhattan

In this section, ANNs was tested against Manhattan's indices of accessibility. The three indices of accessibility were devised again as independent factors (explanatory). The input spatial data was scaled into the range [0 , 1] for both Barcelona and Manhattan [51]. This scaling made these variables compatible with the sigmoid activation function. For this reason, we normalized indices of accessibility using a *lognormal* probability distribution to fall within the range [0 , 1]. The *lognormal* distribution function was chosen because it fits well with the distribution of the three indices [44]. For evaluation, we used the correlation coefficient R^2 to plot block density predictions against actual block density in Manhattan, and we used contingency tables to calculate the ratio of successful scores[5] against misses[6] and false alarms[7].

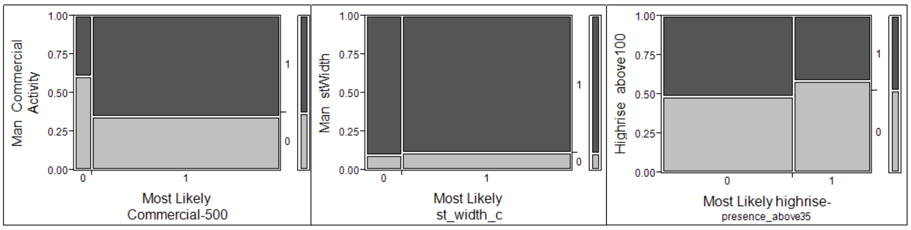
5 Successful scores are where there is an agreement between predicted change and true change

6 Misses are where there are no change predicted but change actually occurred

7 False alarms are where there is change predicted but no change actually occurred

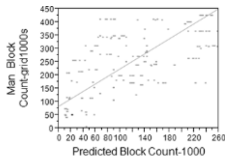
The comparison (Figure 4) showed correspondence in Manhattan. The mosaic and scattergram plots in table 3 showed how response variables correspond to predicted likelihoods. The matching scores between actual data and predicted responses are significant (0.64, 0.74) for commercial activity and street width, but less so for high rise detection. The R^2 value showed a correlation of 0.52 between actual and predicted block density.

Contingency tables for actual parameters by predicted ordinal responses



Commercial activity		Street width	High-rise above 35m
Matching rate[8]	0.65	0.74	0.51

Actual Block count in Manhattan by predicted continuous responses



R^2 = 0.52

Block density per 1000metric square

Fig. 4 Different contingency and scattergram plots elucidating how Manhattan's data corresponds to predictions enabled by ANNs that was initially trained on Barcelona's data.

Forecasting form-function attributes for a hypothetical grid

The validation and testing qualified the ANNs model to be used in forecasting form-function variables for a given spatial structure. This time, the *pixelmapper* method was used to define the approximate features of the urban space. The attributes of the solution space were then defined within that resolution level (Figure 5). The street width response was estimated directly from the NACHslw values and further informed by the ANNs pre-

[8] The rate of true positive and true negative to all scores.

dictions. The rest of the estimated attributes were fully automated assuming a full correspondence between the spatial measures of the winning iteration (iteration 3) and the response variables. The automation was subject to the accuracy of the ANNs model and the scale of representation. Scale might be identified as the metric resolution of the square units in the *pixelmapper* grid. To produce a smooth representation of the target spaces, positive values (1) for ordinal responses were replaced by their correspondent probabilities. Further elaborations on how the pixelated target spaces for the response variables might be translated into 3D descriptions of design solutions was explored in [52].

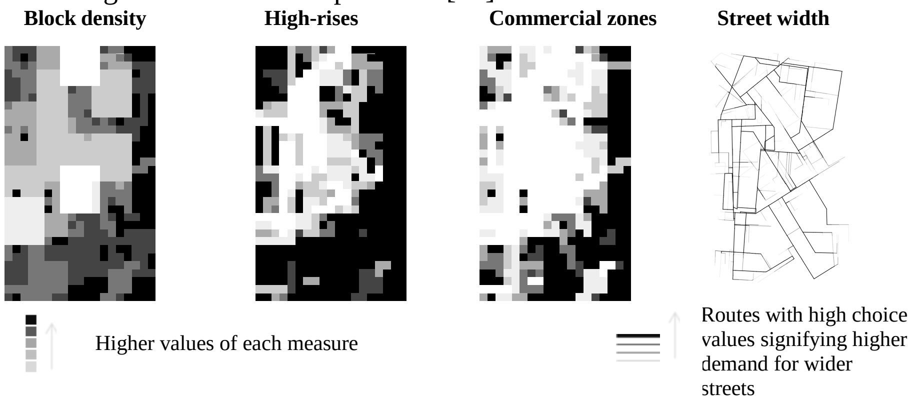


Fig. 5 Responses for form-function estimated by applying the trained and validated ANNs model. The spatial network measures of iteration 3 were used as factors in the ANNs.

Conclusions

The design approach presented here builds and extends on a theoretical urban design model that prioritises the structure of street networks in generating and predicting other features of urban form and function [39], [52], [53]. The theoretical urban design model; namely the prioritised-structure model involved different levels of design filtering starting from the universal; that is a permeable street structure, to define the variables that are thought to be dependent upon street accessibility, based on the assumption that street accessibility has a preference in defining the demand for high rise development, wider streets, dense urban fabric and retail land uses. To encode all these variables we used ANNs.

The design experiment presented in this paper comprised two automated modules. As part of the first module, a generative algorithm was implemented. Four growth iterations were evaluated and compared to a random system and a section of Barcelona's grid structure. The evaluation helped selecting a growth iteration that successfully reproduced the spatial properties witnessed in real cities. The generative process was fully automated. Yet, the evaluation revealed few shortcomings that were either related to the inadequacy of certain measures or to the directional growth mechanisms implemented. Some shortcomings stemmed out of the difficulty to automate a recognition system for certain spatial measures, particularly those related to the definition of street clusters, which might be recognised through spectral analysis [54].

In the forecasting module, ANNs were devised to encode the different topological and geometric measures of street configurations as factors and the different form-function attributes of urban fabric as responses. The model was trained, validated on empirical data from Barcelona, and then tested against data from Manhattan. Data was mapped using an aggregation technique called the *pixelmapper*. The method was introduced in [39], although similar methods were explored in GIS [55]. The *pixelmapper* technique helped binning different types of spatial, binary and continuous data into pixelated square units; hence it was possible to look for invariant relationships in-between different variables within the metric limits of each pixel. Accordingly, a system-based design model was devised using the ANNs activation functions that defined a nonlinear relationship between street network measures and data on form-function in Barcelona. When applied to Manhattan's data, the applicability of ANNs was returned positive. This finding yields with the possibility of applying the functionality of the model on predicting form-function attributes for hypothetical grid structures, hence as a tool to aid urban design. There might be issues, though, that has to do with the computational cost of training ANNs on big data, which might limit the applicability of our proposed model on the regional scale. This effect was trivial in our study, since the largest set of data used in the training was 44093 street segments.

The work presented here encompasses a plausible model to support urban design at the problem definition stage. For a more comprehensive account of the variables that shape urban form, the model needs to be incorporated as part of a broader synthetic model description, considering environmental parameters, and qualitative properties of the urban environment. We only accounted here for variables that might be estimated from street network geometric and topological configurations, where street space acts as a proxy of other urban features. In Space Syntax [2], the affordances of street networks for movement were thought to shape the economic development in cities. This notion was recently recalled in urban morphology [56]. Hence, spatial accessibility is likely to have a preferential role in urban design. Up to this date, space syntax description as an evaluation tool in urban design [57], [20]. In our approach, however, we emphasize that the direct adaptation of analyses into design applications would help supporting design and policy-making practices with empirical evidence through the use of plausible computational models.

REFERENCES

1. Hillier, B, Hanson, J (1984) The social logic of space. Cambridge University Press, Cambridge.
2. Hillier, B. (1996). Space Is The Machine. Cambridge: Cambridge University Press.
3. Ratti, C., Richens, P. (1999). Urban texture analysis with image processing techniques. In Computers in Building (pp. 49-64). Springer US.
4. Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. Proceedings of the National Academy of Sciences, 104(17), 7301-7306.
5. Batty, M., Carvalho, R., Hudson-Smith, A., Milton, R., Smith, D., Steadman, P. (2008). Scaling and allometry in the building geometries of Greater London. The European Physical Journal B, 63(3): 303-314.
6. Barthélemy, M. (2011). Spatial networks. Physics Reports, 499(1), 1-101.
7. Wu, N., Silva, E.A. (2009). Artificial intelligence and ‘waves of complexity’ for urban dynamics. Proc. 8th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering & Data Bases (AIKED '09), 459-464
8. Marshall, S. (2012). Science, pseudo-science and urban design. Urban Design International, 17(4), 257-271.
9. Alexander, C. (1964). Notes on the Synthesis of Form, Cambridge, MA: Harvard University Press.
10. Al_Sayed K., Turner A., Hanna S. (2012). Generative structures in cities, In Proceedings of the 8th International Space Syntax Symposium, Edited by M Greene, J. Reyes, A. Castro, Santiago de Chile: PUC, 2012.
11. Banister, D. Penn, A. Hillier, B. WRG Xu, J (1998) Configurational modelling of urban movement networks. Environment and Planning B: Planning and Design, 25 (1) 59 - 84.
12. Alexiou, K., Johnson, J., Zamenopoulos, T. (Eds.). (2010). Embracing Complexity in Design. London: Routledge.
13. Portugali, J., Meyer, H., Stolk, E., Tan, E. (Eds.). (2012). Complexity theories of cities have come of age: an overview with implications to urban planning and design. Springer.
14. Jacobs, J. (1961). Death and Life of Great American Cities – The Failure of Town Planning, Harmondsworth: Penguin Books.
15. Wu, F. (2002). Complexity and urban simulation: towards a computational laboratory. In Geography Research Forum. Ben-Gurion University of the Negev Press, 22-40.
16. Batty, M. (2010). Networks, flows, and geometry in cities: a challenge to space syntax. The Journal of Space Syntax, 1(2), 366.
17. Ratti, C. (2004). Urban texture and space syntax: some inconsistencies. Environment and Planning B: Planning and Design, 31(4), 487-499.
18. Teeling, C. (1996). Algorithmic Design: Generating Urban Form, , 2, pp. 89- 100.

19. Duarte, J. P., Rocha, J. M., Soares, G. D. (2007). Unveiling the structure of the Marrakech Medina: A shape grammar and an interpreter for generating urban form. AI EDAM, 21(4): 317-349.
20. Duarte, J. P., Beirão, J. N., Montenegro, N., Gil, J. (2012). City Induction: a model for formulating, generating, and evaluating urban designs. In Digital Urban Modeling and Simulation (pp. 73-98). Springer Berlin Heidelberg.
21. Alexander, C., Ishikawa, S., Silverstein, M. (1977). A pattern language: towns, buildings, construction (Vol. 2). Oxford University Press.
22. Stiny, G. (1981) A note on the description of designs. Environment and Planning B: Planning and Design 8, 257–267.
23. Dietzel, C., Herold, M., Hemphill, J. J., Clarke, K. C. (2005). Spatio-temporal dynamics in California's Central Valley: Empirical links to urban theory. International Journal of Geographical Information Science, 19(2): 175-195.
24. Batty, M. (2005). Cities and Complexity: Understanding Cities with Cellular Automata, Agent-based Models, and Fractals. Cambridge, Mass: MIT Press.
25. Coates, P. (2010). Programming Architecture. Routledge.
26. Brown, F. E., Johnson, J. H. (1985). An interactive computer model of urban development: the rules governing the morphology of mediaeval London. Environment and Planning B: Planning and Design, 12(4): 377-400.
27. Parish, Y. I., Müller, P. (2001). Procedural modeling of cities. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques (pp. 301-308). ACM.
28. Derix, C., Gamlesæter, Å., Miranda, P., Helme, L., Kropf, K. (2012). Simulation Heuristics for Urban Design. In Digital Urban Modeling and Simulation (pp. 159-180). Springer Berlin Heidelberg.
29. Stanilov, K., Batty, M. (2011) Exploring the historical determinants of urban growth patterns through cellular automata. Transactions in GIS, 15(3): 253–271, July 2011
30. Marcus, L. (2010). Spatial Capital. The Journal of Space Syntax, 1(1): 30-40.
31. Porta, S., Latora, V., Wang, F., Rueda, S., Strano, E., Scellato, S. Latora, L. (2012). Street centrality and the location of economic activities in Barcelona. Urban Studies, 49(7): 1471-1488.
32. Lämmer, S., Gehlsen, B., Helbing, D. (2006). Scaling laws in the spatial structure of urban road networks. Physica A: Statistical Mechanics and its Applications, 363(1), 89-95.
33. Hillier, B. (2002) A theory of the city as object: Or, how spatial laws mediate the social construction of urban space. Urban Design International 7: 153–179.
34. Zheng, X. Zhao, L. Su, Y. Yan G. Wang, S. (2010) An extension method of Space Syntax and application, Advancing Computing, Communication, Control and Management . Lecture Notes in Electrical Engineering, 56: 7-14.
35. Siksna, A. (1997). The effects of block size and form in North American and Australian city centres. Urban Morphology, 1(1): 19-33.
36. Turner, A (2000) Angular analysis: A method for the quantification of space, Working Paper 23, Centre for Advanced Spatial Analysis, UCL, London
37. Hillier, B., Iida, S. (2005). Network and psychological effects in urban movement. In A. G. Cohn, & D. M. Mark (Eds.), Proceedings of spatial information theory: International conference 2005 (468–476), Ellicottsville, COSIT.

38. Hillier, B., Yang, T. Turner, A. (2012) Normalising least angle choice in Depthmap - and how it opens up new perspectives on the global and local analysis of city space. The Journal of Space Syntax, 3(2), 155-193.
39. Al_Sayed K. (2012) Space Syntax as a parametric model, Proceedings of the Fall 2011 PUARL International Conference, Oregon, Portland.
40. Ozbil A, Peponis J, Stone B, (2011), Understanding the link between street connectivity, land use and pedestrian flows. Urban Design International 16: 125-141
41. Coates, P. (2010). Programming. architecture. Routledge.
42. Jiang, B. (2007). A topological pattern of urban street networks: universality and peculiarity. Physica A: Statistical Mechanics and its Applications, 384(2), 647-655.
43. Al_Sayed K., Turner A., Hanna S. (2010). Modelling the spatial morphogenesis in cities: the dynamics of spatial change in Manhattan, Proceedings of the 10th International Conference on Design And Decision Support Systems In Architecture And Urban Planning, Edited by In: H. JP Timmermans, Eindhoven.
44. Lilliefors, H. W. (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. J. Am. Stat. Assoc. 64:399–402.
45. Conroy, R. (2001). Spatial navigation in immersive virtual environments (Doctoral dissertation, University College London).
46. Turner, A. (2011), UCL Depthmap: Spatial network analysis software, version 10 (London: University College London, VR Centre of the Built Environment, 2011)
47. Openshaw, S., Openshaw, C. (1997). Artificial Intelligence in Geography. Chichester: John Wiley.
48. Rumelhart, D. E., McClelland, J.L., Eds (1986). Parallel distributed processing. Vol.1, Cambridge, MA: MIT Press.
49. Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, 533-536.
50. Foody, G. M. (1996) Relating the land cover composition of mixed pixels to artificial neural network classification output. Photogrammetric Engineering & Remote Sensing, 62: 491–499.
51. Gong, P. (1996) Integrated analysis of spatial data from multiple sources using evidential reasoning and artificial neural-network techniques for geological mapping, Photogrammertic Engineering and Remote Sensing. 62, 513-523.
52. Al_Sayed, K. (2014) Thinking systems in urban design: A prioritised structure model. In Explorations in Urban Design. M. Carmona (ed), (Farnham: Ashgate, 2014), pp. 169–181.
53. Al_Sayed, K. (2014) A systematic approach towards creative urban design. In Design Computing and Cognition DCC'12. J.S. Gero (ed), pp. 133-150. © Springer 2014.
54. Hanna, S. (2012). A representational scheme for the extraction of urban genotypes. In Design Computing and Cognition DCC'12. J.S. Gero (ed). © Springer.
55. Ye, Y., Van Nes, A. (2012). New Ways of Spatial Modelling: Making Spatial Diagnosis in Combining Space Syntax, Spacematrix and MXI with GIS of New and Old Towns. In New Urban Configurations: EAAE/ISUF International Conference. Delft, the Netherlands. October (pp. 16-19).
56. Oliveira V (2013) Morpho, a methodology for assessing urban form, Urban Morphology, 17(1): 149-61.
57. Karimi, K. (2012). A configurational approach to analytical urban design: 'Space Syntax' methodology. Urban Design International, 17(4): 297-318.