

‘IN’ OR ‘AS’ SPACE?: A MODEL OF COMPLEXITY, WITH PHILOSOPHICAL, SIMULATORY, AND EMPIRICAL RAMIFICATIONS

C.H. SMITH

University Libraries, Western Kentucky University, USA.

ABSTRACT

A General Systems model based on ideas originating with the writings of Benedict de Spinoza is described, starting with its philosophical underpinnings and proceeding on to its relation to modern systems concepts, including attempts to simulate the relationships posed and measure real-world structures. Central to the idea is the notion that spatial extension may not have a prior existence but emerges only through an entropy maximization process in which information and energy exchange is balanced among some limited number of subsystems that in sum comprise any given functioning complex system. Related published empiricism concerning geographical/geological systems – the hypsometry of stream basins and overall internal zonation properties of earth structure – is briefly described; the former, especially, reveals a hierarchical pattern of potential energy relations that seems to fit well the organizational hypothesis. Possible applications of the model to genomic codon and medical imaging modeling are alluded to. A brief treatment of the relation of the model to basic properties of complex systems (connectivity, autonomy, emergence, etc.) is provided.

Keywords: complex systems, entropy maximization, General Systems Theory, simulation, space, spatial extension, spatial structure, Spinoza, stream basins.

1 INTRODUCTION

Several years ago, Wilson noted that ‘What characterizes systems of organized complexity is essentially that they are made up of large numbers of parts – and that these parts are strongly connected; that is, they each interact strongly with a number of others’ [1, p. 634]. In a follow-up paper, he went on to suggest that Boltzmann methods might be used to make ‘spatial interaction terms explicit’, thereby giving ‘a new ecological model’ that could provide ‘a potentially more effective way of handling spatial interaction’, and ‘ways of analyzing the dynamics of spatial patterns’ [2, p. 865]. In this work, he mentions that ‘There is only one empirical example in the literature known to the author’, citing an early paper by the present writer [3]. Here I would like to present the outlines of a model relatable to these thoughts, but emerging from a surprisingly different starting point.

As a graduate student in the late 1970s and early 1980s, I took a number of courses in the area of history and philosophy of science. One of these focused on the Vienna Circle school of logical positivism, and especially a famous work by Carnap, *Der Aufbau* [4]. This attempt to generate a logical positivism underpinning for all of science ultimately fell through when Quine [5] observed that Carnap’s system of associating, and relating, objects in space neglected to provide a satisfactory definition for the elemental concept ‘located at’. I began to search for a different way by looking at the concept of interaction in space, and found what I was looking for a couple of years later when I sat in on a class featuring a review of rationalist philosophers, including Benedict de Spinoza (1632–1677).

Spinoza’s assessment of the structure of natural reality, as expressed in his *Ethics*, involved a rather different starting point from the Empiricist approach of Locke and others. Spinoza’s reasoning out of the nature of existence (backed by transcendental proofs that no one seems to understand fully) begins with the posing of what he terms ‘Substance’. ‘Substance’ is removed from any kind of

direct access by ordinary mortals, seemingly representing the very essence of existence (God? or some pantheistic conception thereof?). ‘Substance’ expresses itself through vehicles that Spinoza terms ‘Attributes’, though these are not the directly accessible/measurable attributes of the Empiricist tradition. Instead, they seem to represent some kind of ‘rules of order’ through which the mundane, experienceable, elements of existence (which Spinoza refers to as ‘modifications’ and ‘modes’) are able to express themselves. Spinoza argues that humankind is capable of fathoming only two basic ‘Attributes’: thought and spatial extension. Importantly, each of these independently can provide a complete appreciation of the mundane elements of existence, and in the same basic manner for all of the latter. Recall the old concept of ‘all of nature in a rose’.

It quickly occurred to me that perhaps the secret to such ‘rules of order’ lay at the level of sub-system interactions within each discrete natural system, that is, that perhaps one could divine some basic property or properties of internal interaction (material, energy, and/or information transfer) fundamental to the organization of all such entities. As a first attempt, I set out a model in which natural systems were viewed as subsystemizing according to a ‘lazy universe’ manner of hierarchical organization [6]. This approach was based on an analysis of hierarchical inclusion patterns involving combinatorial mathematics, as shown in Fig. 1. There, one can view how a hierarchy of relations beginning with ‘first-order’ subsystems might self-organize. A model of one of the two basic “Attributes” of Spinoza was thus achieved (this is probably best understood as a model of the Attribute “thought”, though I did not recognize this at first).

It is probably possible to design empirical tests bearing on the validity of this model, but as it represents an understanding of historical differentiation, and thus in practice one-of-a-kind evolutionary processes, it seemed preferable to try to think out a model more amenable to the ecological realities of spatial extension (the other Spinozian attribute). What emerged was a somewhat audacious structure, one which addresses the matter head-on by posing that, perhaps, spatial extension itself represents an entropy maximization process originating with the characteristics of interaction

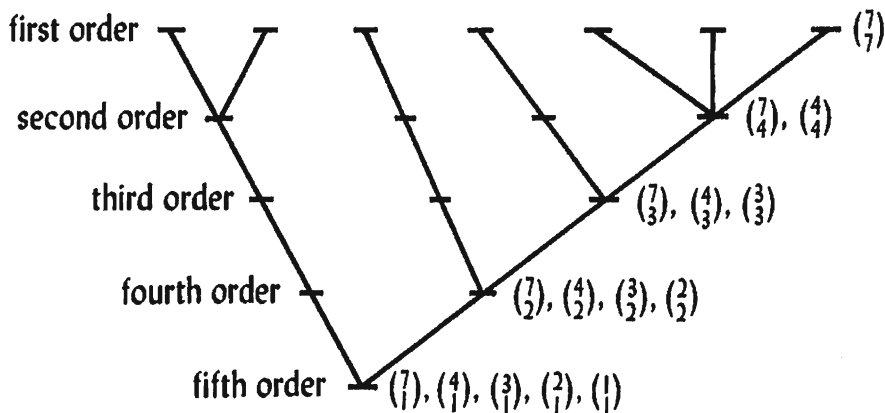


Figure 1: A ‘maximum likelihood tree’ of hierarchical relations constructed with the aid of combinatorial mathematics. Within this structure, all class–subclass relations are ordered on a ‘most-probable-state’ basis. Thus, and for example, the seven first-order subclasses group most probably into one second-order class containing three first-order subclasses, one containing two subclasses, and two containing one subclass. All other relations noted observe the same basic plan. When the structure is extended beyond seven initial elements, it begins to lose this property. (After [6], Fig. 5.)

among a given natural system's (each corresponding to a Spinozian modification) ecological subsystems.

Perhaps the easiest way to imagine this is to think in terms of an $i = j$ – dimensioned input–output matrix depicting a state of standing interaction. A familiar geographical example is the standing pattern of commuter trips among all combinations of towns in some metropolitan region. Assessment of the relative rates of such flows is complicated by the varying populations of the towns involved, but the earlier-mentioned Wilson got past this difficulty by applying entropy maximization methods to the problem [7]. A ‘most-probable-state’ solution to the matrix of interactions was thus achieved which accounted for a large proportion of the variation present in the original flows matrix; once this effect was removed, further causalities (often related to the relative distances among the places involved) could be investigated.

This approach is still in use. To a certain extent, it has lost its appeal as a modeling device, however, devolving into a methodological tool with apparent limited relevance to how urban/regional structures actually *evolve*. The problem, it seems to me, lies in restricting the measurement of flows to arbitrarily defined sources and destinations that do not necessarily reflect the manner of organizational differentiation within natural systems. The papers of Wilson [1] and Wilson *et al.* [2] address this issue.

If we could identify a form of subsystemization that is basic to the structure involved, and whose operation can be expressed in the form of measurable interactions among the elements, we might be better able to get at the relation of system change to its ongoing maintenance of ecological integrity. I conceived of a system structure characterized by a small number of subsystems: (1) that in sum constitute the whole of the system; (2) that share information, energy, and/or materials with one another; (3) in a manner that can be objectively represented as ‘flows’ (implied or actual) in an input–output matrix. Finally, and most interestingly, I think, I propose that an entropy maximization of this kind of intra-system structure might reveal a pattern of symmetry interpretable as three-dimensional space. Otherwise put, I introduce the possibility that it is a continuing entropy maximization of this kind of relations that *is* spatial extension, to the extent that everything that is spatially extended must observe these conditions.

This solution, if it be a real one, is one which seemingly remains faithful to the Spinozian model, while lending itself to examination as something more than just a philosophical structure. In particular, it appears to be particularly amenable to simulation and, indeed, to actual empirical investigations.

2 SIMULATIONS

In attempting to simulate such a set of relations, the most direct approach involves brute force methods. Basically, are there any sets of numbers, expressed in dimensionality $i = j$ matrix form, that when entropy-maximized will produce the anticipated results? If there are none, of course, the thought process here is for naught; conversely, if all sets tested yield the necessary outcome, the model produces, at best, trivial results.

The specific question being investigated in the simulations is whether a given matrix of numbers, when entropy-maximized, will produce results interpretable as a Euclidean space – that is, which can be projected as a set of Euclidean dimensional relations. To examine this, I filled hundreds of thousands of matrices, of dimensions 3×3 , 4×4 , 5×5 , and 6×6 , with six-digit random numbers, and submitted these to an entropy maximization operation known as double-standardization (or bistochastization), in which the rows and columns of the matrix are alternately converted to z-scores for as many times as it takes to converge on stable values. It turns out that only one kind of result directly corresponds to (i.e. ‘produces a projection of’) a set of relations recognizable as Euclidean spatial extension: for a value of $n = 4$ matrix dimensions, $ij = ji$ scores that are symmetric about the

main diagonal. (This can be confirmed through metric multidimensional scaling, the result being a zero-stress solution wherein all the unrotated three-dimensional coordinate points are effectively indistinguishable from one another.) For example, this matrix of raw numbers:

2835	1962	1293	908
1962	2287	2098	1689
1293	2098	2500	2442
908	1689	2442	3615

when double-standardized, produces the z-scores:

1.4562	0.2296	-1.2934	-0.3924
0.2296	1.4562	-0.3924	-1.2934
-1.2934	-0.3924	1.4562	0.2296
-0.3924	-1.2934	0.2296	1.4562

These scores, when interpreted as similarities and transformed into distances, correspond to relations within an unambiguously defined three-dimensional space.

Depending on the exact initial conditions imposed on the simulations, the percentage of $n = 4$ matrices that achieve this symmetric form ranges between about one-half of 1% and 2%. Figure 2 gives a summary of the most important of these results, combining several pieces of information. For each matrix dimensionality tested (again, 3×3 through 6×6), four sets of simulations were run.

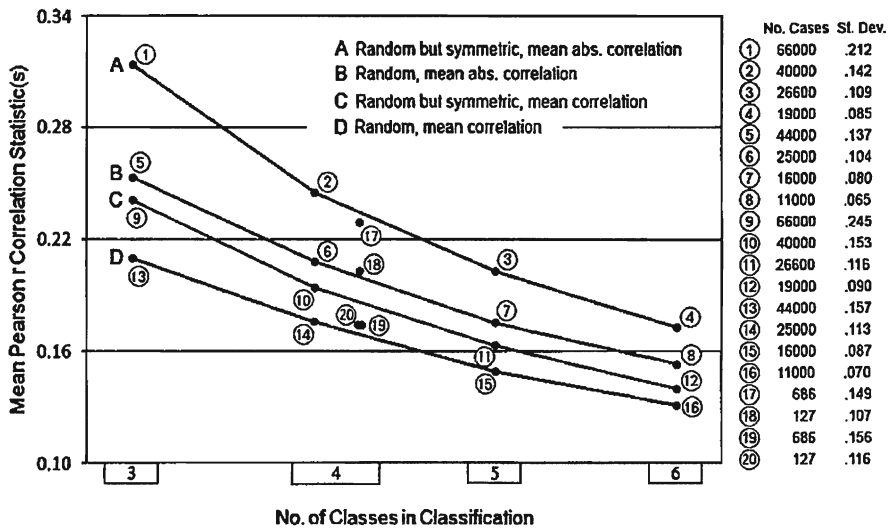


Figure 2: Summary of spatial projection simulations for random numbers matrix configurations of dimension 3×3 through 6×6 . Circled numbers refer to data at the right margin giving the number of simulations in each test and the standard deviations accompanying the mean values plotted. The plotted values are the means obtained for each set of simulations performed at each matrix dimensionality (=‘no. of classes in classification’). Lines connecting the points are for readability purposes only. Point values 17 through 20 are compiled from subsets of the data leading to point values 2, 6, 10, and 14, respectively. See text for further explanation. (After [8], Fig. 1.)

combining two different conditions of random number placement/inclusion (fully random, or random but symmetric about the $i = j$ diagonal) with two secondary statistical measures of the inputted data (mean value of the r values from the associated Pearson r correlation matrix or mean value of the absolute values of the column means of same). Points 17 through 20 in Fig. 2 represent the parallel data for z-score results that passed the $i = j$ symmetry requirement just described; again, these occurred under the $n = 4$ dimensions condition (and thus represent a special subset of the data leading to points 2, 6, 10, and 14). The standard deviations associated with the 20 mean values plotted are rather large, but it is apparent that, broadly speaking, the mean r values decrease with increasing numbers of classes. The significance of this fact will become apparent in a moment.

The results displayed in Fig. 2 make sense in several crucial respects. First, the fact that 17, 18, 19, and 20 exist at all signifies that a small, but not trivially small, number of appropriate solutions exist under the universe of all possible matrix contents (i.e. simulated interaction states). Second, these findings are sympathetic to the fact that four nodes are required to define a three-dimensional Euclidean space. Third, the high standard deviations connected with each mean plotted as 17, 18, 19, and 20 above indicate a wide range of intra-correlations among the vectors of the matrices involved, a potentially important clue to the way systems may evolve internally toward reduction of redundancies of function through feedbacks. Fourth, and a related consideration, equations representing the functions of systems involving five or more variables are almost never solvable, making it possible to imagine how a four-subsystem structure might self-adjust.

The success of the random numbers simulations suggested an elaboration. So, soon after completing them, I began to investigate point patterns (spaced regularly on geometric surfaces) grouped randomly, or arbitrarily, into class structures (here, of four classes). In this instance, the summary values filling the raw data matrices were metric spatial autocorrelation scores (which provided a measured indication of the relative overall location of the members of each class of 'sampled' points with respect to the overall locations of the members of each other class). These studies are described elsewhere [8–10]. Briefly, I found that roughly 5% of the systems spreading out over arbitrarily bounded two-dimensional surfaces passed the symmetrical z-scores test. This percentage increased further when several configurations within, and across the surface of, spherical systems were analyzed. In all, an 'evolutionary' sequence is suggested in which it becomes easier and easier to achieve a symmetric z-scores result, as the prior geometry changes from undefined, to two-dimensional, to three-dimensional.

3 REAL-WORLD SYSTEMS

The success of the simulations provided an impetus to continue on to empirical analyses of actual natural systems. A number of pilot studies on systems ranging from butterfly wing color patterns to regional human populations were attempted, with some success [8–10]. Here I would like to describe the two most successful of these, which produced published works [8,9].

In this context, it would be ideal to investigate systems characterized by actual and discrete energy, materials, or information flows that can be directly measured, but the body of knowledge that could sustain such measurements may not exist as yet. The next best thing seems to be an investigation of patterns which represent 'implied flows', that is, the structural result, in space, of continuing processes. For good examples of these, we may turn to topographical and geological models.

Stream basins represent bounded systems characterized by relatively well-worked out topographical patterns that are due largely to a simple universal driving mechanism: gravity. Thus, every point on the surface of any given basin has a certain potential energy connected with it, and ultimately it is the working out of these potentials with respect to one another that leaves us with the topography witnessed at any given time. The question posed here is whether such interactions operate as

singularities (as do the commuter trips between a large number of origins and destinations) or actually respond to causal forces operating primarily at a grouped level (most likely, as ranges of elevations).

As described in more detail in the published account, the elevations within 31 similarly-sized drainage basins in the Commonwealth of Kentucky, USA, were sampled at three levels of spatial fineness to produce three sets of data; each of these were then assembled into rank-ordered vectors. These data were then subjected to a series of nonhierarchical, information statistic-based, cluster analyses, classifying the contents of the 93 vectors into 3-, 4-, 5- and 6-class solutions.

The cluster analyses produced unexceptional results diagrammed as Fig. 2 of [8]; at all three fineness levels of sampling, the variations explained increased, on the average, in a very regular pattern according to the number of classes imposed. But this analysis did not take into account the relative spatial positions of all the point elevations sampled, and when this was done, a very different picture emerged.

All 93×4 sets of data were subjected to a pair of metric spatial autocorrelation analyses, that is, the spatial relations of the sampled points within each range of elevations were related, as a group, to the relative spatial positions of the points within the other groups (and to themselves). This produced $93 \times 4 \times 2$ matrices of spatial autocorrelation coefficients (of dimensions 3×3 , 4×4 , 5×5 , and 6×6), which were then double-standardized as described in the simulations section.

Of the results, as was expected, none of the operations on the three-, five-, and six-class structures produced the symmetric z-scores pattern described earlier. For the four-class solutions, 28 of 31 basins passed the test at the finest sampling levels, but only 18 of 31 did at the coarsest one (the real-world distance between nearest neighbors in the finest sampling grid was 215 feet, the second finest 430 feet, and the coarsest 860 feet). A special, even finer, grid was applied to sample the three basins that did not pass the test, whereupon two of the three then passed the test, with the third showing considerable movement toward passing, but not quite doing so.

A perhaps even more interesting result is depicted in Fig. 3. This chart is organized in a manner similar to Fig. 2, but instead of displaying the results of simulation exercises, it summarizes the stream basins analysis. All of the values plotted represent means across each set of 31 results (the numbers next to each datum point are the accompanying standard deviations). Two sets of results are presented, based on the two internal redundancy measures employed. Within each set, the results for each classification model at the three different finenesses of sampling are shown.

In contrast with Fig. 2, the trend toward decreasing mean r values with increasing number of class divisions is broken up abruptly at the four-class level; this is apparent in all six sets of models and is thus a very robust result. In fact, in all six sets of models, the *lowest* mean r value is associated with the four-class model. It thus seems apparent that *something* is producing a special causal effect at this level.

Again, the significance of the mean r statistic is that it possibly provides a glimpse into the organization of redundancies within a system – at least to those operating at the subsystem level as defined here. The lowest mean r possible for a correlation coefficient matrix is zero; under such conditions, each vector of values in the matrix is as different as is mathematically possible from the other vectors, taken as a group. It is tempting in the present instance to regard real-world examples exhibiting rather high mean r scores as representing systems that are unstable, that is, as operating under the influence of feedbacks that are serving to move the entity toward higher levels of efficient operation. The results of the analysis discussed here seem to fortify this impression, as the three-, five-, and six-class models show no evidence of any trend beyond that of a simple statistical nature (similar to the increasing ‘variation explained’ cluster analysis results discussed earlier), whereas the four-class model does.

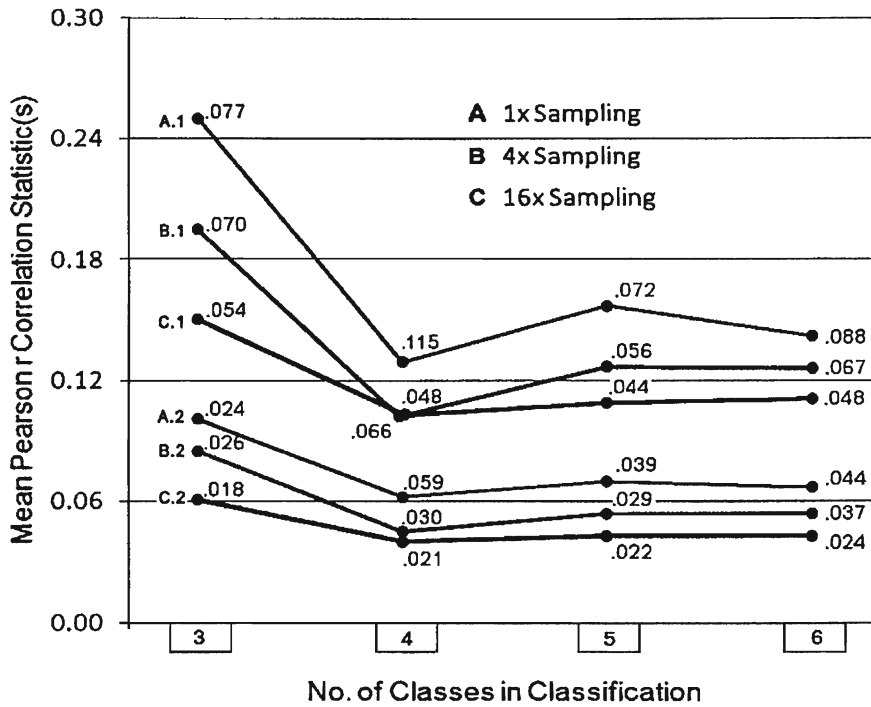


Figure 3: Summary of spatial subsystemization properties of topography in 31 Kentucky stream basins, based on three fineness levels of sampling and applying the two mean correlation matrix formulations (top set of three, mean of the absolute values; bottom set of three, mean of the means). The associated standard deviation for each mean value plotted is written out next to it. The lines connecting the means plotted are for readability purposes only. See text for further explanation. (After [8], Fig. 3.)

One may reasonably ask whether the results displayed here might represent simple reconstructions of the three-dimensionality already built into the setting, but three arguments can be presented to show that this is an unlikely interpretation. First, the mean r values for the four-class solutions here are much lower than for the parallel simulations noted earlier. Second, a follow-up analysis was performed on the 1x and 16x samplings in which the vectors of elevations were arranged into unnatural (but still range-of-values-related) class structures; when this was done and the resulting spatial autocorrelation statistic-based matrices were entropy-maximized, only a small percentage of the models passed the z-scores symmetry test, and the resulting mean r values approached those from the simulations. Third, in an earlier pilot study involving 25 stream basins, results consistent with (but not so striking as, being based on coarser data) the ones discussed here were obtained, while a different kind of follow-up analysis was administered. It was reasoned that those individual basins that yielded high mean r values should be associated with topographical basin conditions indicative of imbalances between internal rates of erosion and deposition: in short, that the slopes across such basins should be less regular. A fairly crude measure of same was applied, with the result that nearly 50% of the variation among the 25 mean r values was explained by this variable alone [10].

The second analysis, published in 2014 [9], may now be described briefly. It also involved a spatial autocorrelation-based approach. Many natural systems, as interpreted through the model

discussed here, might be expected to exhibit rather high internal redundancies, either as a result of natural development trends or in temporary response to external influences (as would be the case with the stream systems noted above). I wanted to investigate some system that one would expect to have highly settled internal relations, the results of a very long and relatively undisturbed evolution. The internal zones of the earth seemed to fit this description; so, I have undertaken several analyses of the spatial relations among the inner core, outer core, mantle, and 'surface zones' (oceanic and continental crust, plus hydrosphere and atmosphere). The results of these studies may be found in [9, 10]. Earlier simulation studies [10] on the development of concentric zones with spheres and spheroids had determined that depending on the exact measures applied, a greatly varying proportion of the models passed the z-scores symmetry test. The results of the analysis of the actual pattern of internal zonation of the earth passed the symmetry test and produced mean r scores of as low as 0.001, a figure so close to zero as to be surpassed in this respect only a handful of times across the hundreds of thousands of random numbers simulations I have performed. In the latest analysis [9], mean r values of 0.003 were obtained, not only for the present configuration of continents and oceans, but also for two other periods in geological history extending backward some 200 million years. These were compared to an idealized earth structure model in which the 'surface zones' were averaged out to create a zone of consistently equal depth, in contrast with actual conditions (in which continental crusts, especially, cause a zone of greatly varying depths around the world). The idealized structure also passed the symmetry test, but the mean r values, though still very small, were about twice those connected with the actual earth patterns. These results are quite suggestive, but will require a sampling grid of points within the earth much finer than I was able to apply (involving between 15,000 and 20,000 internal earth locations) to fully settle the matter.

4 DISCUSSION

My efforts at developing this model, continued over a more than 30-year period, have been slowed by a number of factors, not the least of which is its generally 'foreign' feel. Yet, its investigation does not seem to violate any basic tenets of the methodology of science, and so far, the results obtained have not been sufficient to make me think that efforts in this direction should be shut down. Some of these results, especially from the stream basins work, appear difficult to dismiss that easily.

Assuming for the moment that this Spinoza-inspired understanding of the way complex systems are organized is actually (at least, to an approximation) correct, a number of very important applications can be imagined. Philosophically, of course, this is a model of nature invoking final causes. But the final causes imagined are quite distinct from the old teleology, dependent on notions of supra-normal intervention toward predestined ends. Instead, systems are imagined to emerge as temporary spatial structures from their antecedents, only to exhaust their potential at some point and be re-absorbed by their surroundings (in the case of living things, 'die'). Their organization may be based on prior programs (as in the case of DNA-based life), but within that program there is the flexibility to change, so long as that change results in novel structure still observing the fundamental constraints on subsystemization suggested here.

As an evolutionary model, therefore, this understanding is distinct from Darwinian doctrine in steering away from anything describable in simple 'random walk' terms. Yes, an infinite number of potentialities of development remain in any given instance, but a 'much smaller' infinite number than we now assume to be the case.

As a practical matter, it must be remembered that to hold water, the model here must pertain to literally everything that can be construed as a naturally developing structure (including elements of human civilization, though not extending so far as providing an ultimate meaning for vacuum cleaners!). This means not only at the mesoscale we are most familiar with, but also with the very

large and very small. It is perhaps not irrelevant, for example, that the fundamental interactive forces of physics and the codon genetics of the DNA molecule are each connected to elemental phenomena that are four in number.

For the present, it seems incumbent to develop one or two applications that will really serve to focus attention. Stream basins, though perhaps an ideal starting point from the point of view of measurement, will ultimately not accomplish this; a more detailed look at the spatial configuration of the earth's internal zones might serve better. But best of all would be some medical application, possibly as related to medical imaging. The human brain, for example, exhibits a lot of properties that are not dissimilar to the potential energies present across a drainage basin, and might be measured in analogous fashion. Brain temperature, electrical activity, blood flow rates, etc. are general properties all of whose measures could be sampled spatially in just the fashion I have described here, and subjected to analyses designed to expose their patterns of operation. It would be supposed that a fully healthy brain, analyzed to the ends I have discussed, should produce a final mean r measure little deviating from zero. Hypothetically, the more diseased and/or degenerating the structure, the more it should measure out to produce a mean r statistic of higher value. Concerted study in this direction might make it possible to identify, for example, susceptibility to, or early onset of, Alzheimer's disease. A research strategy in this direction is easy to imagine.

Possible applications of the 'maximum likelihood tree' model are less easy to imagine, but even there such an approach might prove useful to understanding, say, the pattern of ontological development of organ systems in development, or even to the pattern of ongoing emergences of ocean crust plates in plate tectonics.

I do not claim that the materials I have presented here and elsewhere prove the validity of this model. But, in fact, even if subsequent studies do not bear it out, the more basic idea of a 'final causes' approach toward complex systems it suggests should also provoke thought, as it introduces a way of understanding how system-level constraints on 'random walk' processes may operate.

REFERENCES

- [1] Wilson, A.G., Ecological and urban systems models: some explorations of similarities in the context of complexity theory. *Environment and Planning A*, **38**, pp. 633–646. 2006.
- [2] Wilson, A.G., Boltzmann, Lotka and Volterra and spatial structural evolution: an integrated methodology for some dynamical systems. *Journal of the Royal Society Interface*, **5**, pp. 865–871.
- [3] Smith, C.H., A system of world mammal faunal regions. II. The distance decay effect upon inter-regional affinities. *Journal of Biogeography*, **10(5)**, pp. 467–482.
- [4] Carnap, R., *Der Logische Aufbau der Welt*. Weltkreis-Verlag: Berlin-Schlachtensee, 1928.
- [5] Quine, W.V., Two dogmas of empiricism. *Philosophical Review*, **60**, pp. 20–43, 1951.
- [6] Smith, C.H., A general approach to the study of spatial systems. I. The relational representation of measurable attributes. *International Journal of General Systems*, **12(4)**, pp. 359–384, 1986.
- [7] Wilson, A.G., *Entropy in Urban Planning*. Pion: London.
- [8] Smith, C.H. & Derr, M., 'In space' or 'as space'? a new model. *Life*, **2(3)**, pp. 243–254, 2012.
- [9] Smith, C.H., 'In space' or 'as space'? spatial autocorrelation properties of the earth's interior. *International Journal of Geosciences*, **5(4)**, pp. 375–382, 2014.
- [10] The Once and Future Wallace: A Theory of Spatial Systems, Online. <http://people.wku.edu/charles.smith/once/writings.htm#2>