IntechOpen

Theory of Complexity
Definitions, Models, and Applications

Edited by Ricardo López-Ruiz

Theory of Complexity - Definitions, Models, and Applications

Edited by Ricardo López-Ruiz

Published in London, United Kingdom



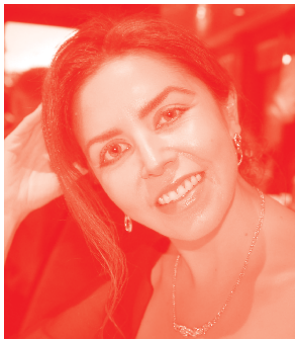




IntechOpen





Supporting open minds since 2005

Theory of Complexity - Definitions, Models, and Applications
http://dx.doi.org/10.5772/intechopen.83237
Edited by Ricardo López-Ruiz

Contributors
Heidi Kloos, Michael J. Droboniku, Dieter Vanderelst, Blair Eberhart, Weiqi Li, José Roberto Castilho Piqueira, Sérgio Henrique Vannucchi Leme de Mattos, Luiz Eduardo Vicente, Andrea Koga Vicente, Claudio Bieleni Junior, Maristella Cruz de Moraes, Gabriele Luiza Cordeiro, Sergio Velázquez-Medina, Ulises Portero-Ajenjo, Rafael Prieto Curiel

First published in London, United Kingdom, 2021 by IntechOpen
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 5 Princes Gate Court, London, SW7 2QJ, United Kingdom
Printed in Croatia

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Additional hard and PDF copies can be obtained from orders@intechopen.com

Theory of Complexity - Definitions, Models, and Applications
Edited by Ricardo López-Ruiz
p. cm.
Print ISBN 978-1-78985-213-4
Online ISBN 978-1-78985-214-1
eBook (PDF) ISBN 978-1-83962-855-9

Meet the editor

Ricardo López-Ruiz, MS, Ph.D., is an associate professor in the Department of Computer Science and Systems Engineering, Faculty of Science, University of Zaragoza, Spain. He is also an associate researcher in Complex Systems at the School of Mathematics, University of Zaragoza. Previously, he worked as a lecturer at the University of Navarra, the Public University of Navarra, and the UNED of Calatayud. He completed his postdoc with Prof. Yves Pomeau at the École Normale Supérieure, Paris, and with Prof. Gabriel Mindlin at the University of Buenos Aires. His areas of interest include statistical complexity and nonlinear models, chaotic maps and applications, multiagent systems, econophysics, big data, and machine learning techniques.

Contents

Preface

The meaning of complexity is a subject of intense debate in exact sciences, systems biology, and social science. Systems of many components, such as proteins at the cellular level, cells at the organ level, or agents at the social level, provide complexity science a ubiquity that has percolated all branches of knowledge. In all these cases, emergent behaviors result from the coordinated action of different components and constitute new global system properties. These new properties allow the system to adapt to the environment in an organized manner that will continue to demand new features for its survival. This leads to the evolution of systems as a natural consequence of the complexity associated with them.

This book intends to deepen the general meaning of complexity from different points of view, to inquire about the different statistical and computational valid paradigms in social and natural systems. The first part focuses on social systems and includes chapters on different approaches to the traveling salesman problem (Chapter 1 by Weiqi Li), models of opinion dynamics creation (Chapter 2 by Prieto Curiel), and a universal theory for knowledge formation in children (Chapter 3 by Droboniku, Kloos, et al.). The second part addresses different natural systems from a complexity perspective, in particular, the evaluation of landscape organization and dynamics through information entropy indicators (Chapter 4 by Piqueira et al.) and the study of the performance of wind farms with the use of artificial neural networks (Chapter 5 by Velázquez-Medina and Portero-Ajenjo). All these chapters present some new perspectives and applications within the broad field of complexity science. We hope that this book will be useful as a guide to an audience interested in the different problems and approaches within the theory of complexity.

As the editor of this book, I would like to thank all the authors who have contributed to this volume. I must also express my gratitude to the editorial staff at IntechOpen, particularly Author Service Manager Ms. Marijana Francetic. I am also grateful to my family, friends, and advisors. Finally, I would like to dedicate this book to the victims of COVID-19.

Ricardo López-Ruiz
University of Zaragoza,
Spain

Section 1

Theory of Complexity in Social Systems

Chapter 1

How to Solve the Traveling Salesman Problem

Weiqi Li

Abstract

The Traveling Salesman Problem (TSP) is believed to be an intractable problem and have no practically efficient algorithm to solve it. The intrinsic difficulty of the TSP is associated with the combinatorial explosion of potential solutions in the solution space. When a TSP instance is large, the number of possible solutions in the solution space is so large as to forbid an exhaustive search for the optimal solutions. The seemingly "limitless" increase of computational power will not resolve its genuine intractability. Do we need to explore all the possibilities in the solution space to find the optimal solutions? This chapter offers a novel perspective trying to overcome the combinatorial complexity of the TSP. When we design an algorithm to solve an optimization problem, we usually ask the critical question: "How can we find all exact optimal solutions and how do we know that they are optimal in the solution space?" This chapter introduces the Attractor-Based Search System (ABSS) that is specifically designed for the TSP. This chapter explains how the ABSS answer this critical question. The computing complexity of the ABSS is also discussed.

Keywords: combinatorial optimization, global optimization, heuristic local search, computational complexity, traveling salesman problem, multimodal optimization, dynamical systems, attractor

1. Introduction

The TSP is one of the most intensively investigated optimization problems and often treated as the prototypical combinatorial optimization problem that has provided much motivation for design of new search algorithms, development of complexity theory, and analysis of solution space and search space [1, 2]. The TSP is defined as a complete graph $Q = (V, E, C)$, where $V = \{v_i : i = 1, 2, \ldots, n\}$ is a set of n nodes, $E = \{e(i,j) : i, j = 1, 2, \ldots, n; i \neq j\}_{n \times n}$ is an edge matrix containing the set of edges that connects the n nodes, and $C = \{c(i,j) : i, j = 1, 2, \ldots, n; i \neq j\}_{n \times n}$ is a cost matrix holding a set of traveling costs associated with the set of edges. The solution space S contains a finite set of all feasible tours that a salesman may traverse. A tour $s \in S$ is a closed route that visits every node exactly once and returns to the starting node at the end. Like many real-world optimization problems, the TSP is inherently multimodal; that is, it may contain multiple optimal tours in its solution space. We assume that a TSP instance Q contains h (≥ 1) optimal tours in S. We denote $f(s)$ as the objective function, $s^* = \min_{s \in S} f(s)$ as an optimal tour and S^* as the set of h optimal tours. The objective of the TSP is to find all h optimal tours in the solution space, that is, $S^* \subset S$. Therefore, the argument is

$$\arg\left[\min_{s\in S} f(s)\right] = S^* = \left[s_1^*, s_2^*, \ldots, s_h^*\right] \quad (1)$$

Under this definition, the salesman wants to know what all best alternative tours are available. Finding all optimal solutions is the essential requirement for an optimization search algorithm. In practice, knowledge of multiple optimal solutions is extremely helpful, providing the decision-maker with multiple options, especially when the sensitivity of the objective function to small changes in its variables may be different at the alternative optimal points. Obviously, this TSP definition is elegantly simple but full of challenge to the optimization researchers and practitioners.

Optimization has been a fundamental tool in all scientific and engineering areas. The goal of optimization is to find the best set of the admissible conditions to achieve our objective in our decision-making process. Therefore, the fundamental requirement for an optimization search algorithm is to find *all optimal solutions* within a *reasonable amount of computing time*. The focus of computational complexity theory is to analyze the intrinsic difficulty of an optimization problem and the asymptotic property of a search algorithm to solve it. The complexity theory attempts to address this question: "How efficient is a search algorithm for a particular optimization problem, as the number of variables gets large?"

The TSP is known to be NP-hard [2, 3]. The problems in NP-hard class are said to be intractable because these problems have no asymptotically efficient algorithm, even the seemingly "limitless" increase of computational power will not resolve their genuine intractability. The intrinsic difficulty of the TSP is that the solution space increases exponentially as the problem size increases, which makes the exhaustive search infeasible. When a TSP instance is large, the number of possible tours in the solution space is so large to forbid an exhaustive search for the optimal tours. A feasible search algorithm for the TSP is one that comes with a guarantee to find all best tours in time at most proportional to n^k for some power k.

Do we need to explore all the possibilities in the solution space to find the optimal solutions? Imagine that searching for the optimal solution in the solution space is like treasure hunting. We are trying to hunt for a hidden treasure in the whole world. If we are "blindfolded" without any guidance, it is a silly idea to search every single square inch of the extremely large space. We may have to perform a random search process, which is usually not effective. However, if we are able to use various clues to locate the small village where the treasure was placed, we will then directly go to that village and search every corner of the village to find the hidden treasure. The philosophy behind this treasure-hunting case for optimization is that: if we do not know where the optimal point is in the solution space, we can try to identify the small region that contains the optimal point and then search that small region thoroughly to find that optimal point.

Optimization researchers have developed many optimization algorithms to solve the TSP. Deterministic approaches such as exhaustive enumeration and branch-and-bound can find exact optimal solutions, but they are very expensive from the computational point of view. Stochastic optimization algorithms, such as simple heuristic local search, Evolutionary Algorithms, Particle Swarm Optimization and many other metaheuristics, can find hopefully a good solution to the TSP [1, 4–7]. The stochastic search algorithms trade in guaranteed correctness of the optimal solution for a shorter computing time. In practice, most stochastic search algorithms are based on the heuristic local search technique [8]. Heuristics are functions that help us decide which one of a set of possible solutions is to be selected next [9]. A local search algorithm iteratively explores the neighborhoods of solutions trying to improve the current solution by a local change. However, the scope of local search is

limited by the neighborhood definition. Therefore, heuristic local search algorithms are locally convergent. The final solution may deviate from the optimal solution. Such a final solution is called a *locally optimal solution*, denoted as s' in this chapter. To distinguish from locally optimal solutions, the optimal solution s^* in the solution space is usually called the *globally optimal solution*.

This chapter studies the TSP from a novel perspective and presents a new search algorithm for the TSP. This chapter is organized in the following sections. Section 2 presents the ABSS algorithm for the TSP. Section 3 describes the important data structure that is a critical player in solving the TSP. Section 4 discusses the nature of heuristic local search algorithm and introduces the concept of solution attractor. Section 5 describes the global optimization features of the ABSS. Section 6 discusses the computational complexity of the ABSS. Section 7 concludes this chapter.

2. The attractor-based search system for the TSP

Figure 1 presents the Attractor-Based Search System (ABSS) for the TSP. In this algorithm, Q is a TSP instance with the edge matrix E and cost matrix C. At beginning of search, the matrix E is initialized by assigning zeros to all elements of E. The function `InitialTour()` constructs an initial tour s_i using any tour-construction technique. The function `LocalSearch()` takes s_i as an input, performs local search using any type of local search technique, and returns a locally optimal tour s_j. The function `UpdateE()` updates the matrix E by recording the edge configuration of tour s_j into the matrix. K is the number of search trajectories. After the edge configurations of K locally optimal tours are stored in the matrix E, the function `ExhaustedSearch()` searches E completely using the depth-first tree search technique, which is a simple recursive search method that traverses a directed graph starting from a node and then searches adjacent nodes recursively. Finally, the ABSS outputs a set of all best tours S^* found in the edge configuration of E. The search strategy in the ABSS is straightforward: generating K locally optimal tours, storing their edge configurations in the matrix E, and then identifying the best tours by evaluating all tours represented by the edge configuration of E. The ABSS is a simple and efficient computer program that can solve the TSP effectively. This search algorithm shows strong features of effectiveness, flexibility, adaptability, scalability and efficiency. The computational model of the ABSS is inherently parallel, facilitating implementation on concurrent processors. It can be implemented in many different ways: series, parallel, distributed, or hybrid.

```
1   ABSS(Q)
2   begin
3     Initialize E;
4     NumberOfTrajectories = 0;
5     repeat
6       s_i = InitialTour();
7       s_j = LocalSearch(s_i);
8       E = UpdateE(s_j, E);
9       NumberOfTrajectories = NumberOfTrajectories + 1;
10    until NumberOfTrajectories = K
11    S* = ExhaustedSearch(E);
12  end
```

Figure 1.
The ABSS algorithm for the TSP.

Figure 2 uses a 10-node instance as an example to illustrate how the ABSS works. We randomly generate $K = 6n = 60$ initial tours, which edge configurations hit all elements of the matrix E (marked as black color), as shown in **Figure 2(a)**. It means that these 60 random tours hit all 45 edges that represent all 181440 tours in the solution space. We let each of the search trajectories run 5000 iterations and obtain 60 locally optimal tours. However, due to the small size of the instance, most locally optimal tours have identical edge configurations. Among the 60 locally optimal tours, we find only four distinct locally optimal tours as shown in **Figure 2(b)**. **Figure 2(c)** shows the union of the edge configurations of the 60 locally optimal tours, in which 18 edges are hit. Then we use the depth-first tree search, as illustrated in **Figure 2(d)**, to identify all five tours in the edge configuration of E, which are listed in **Figure 2(e)**. In fact, one of the five tours is the globally optimal tour. This simple example indicates that (1) local search trajectories converge to small set of edges, and (2) the union of the edge configurations of K locally optimal tours is not just a countable union of the edge configurations of the these tours, but also include the edge configurations of other locally optimal tours. The ABSS consists of two search phases: local search phase and exhaustive search phase.

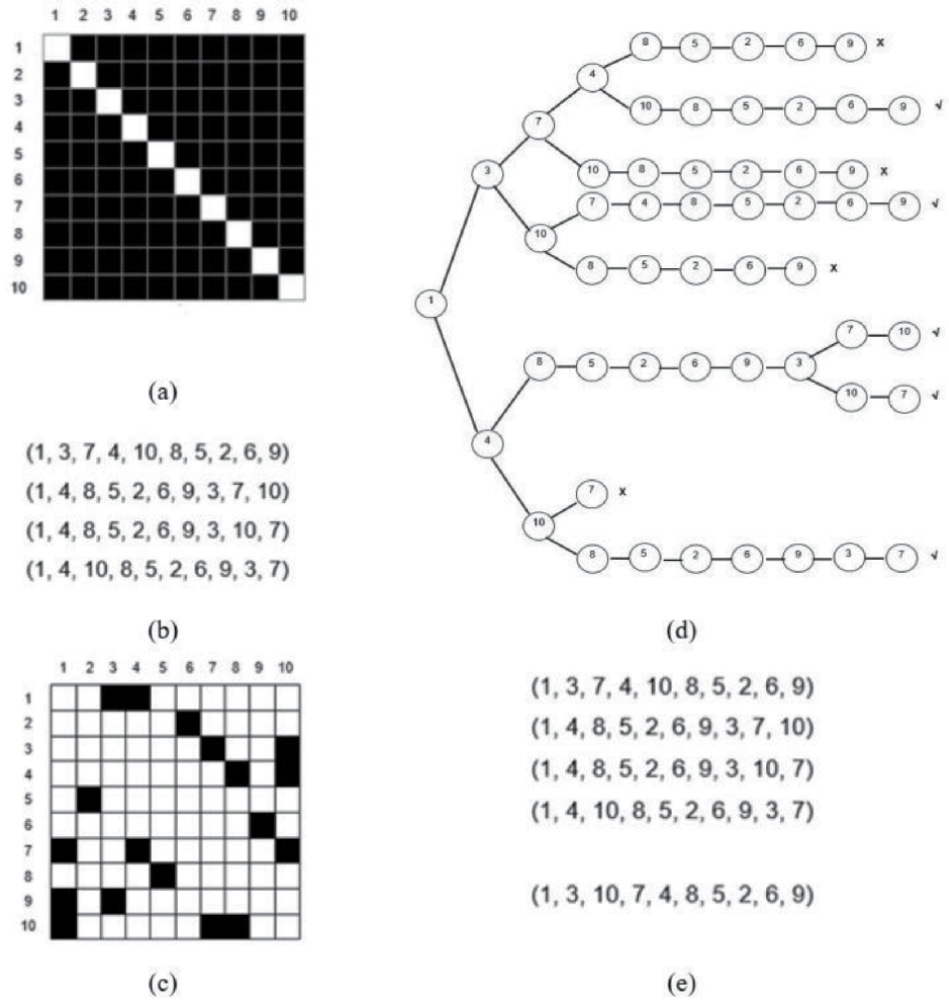


Figure 2.
A simple example of the ABSS algorithm. (a) Union of the edge configurations of 60 random initial tours, (b) four distinct locally optimal tours, (c) union of the edge configurations of the 60 locally optimal tours, (d) the depth-first tree search on the edge configuration of E*, and (e) five tours found in* E*.*

The task of the local search phase is to identify the region that globally optimal tour is located (i.e. the village hiding the treasure), and the task of the exhaustive search phase is to find the globally optimal tour (i.e. find the hidden treasure). The remaining sections will briefly explain the features of the ABSS.

In all experiments mentioned in the chapter, we generate symmetric TSP instances with n nodes. The element $c(i,j)$ of the cost matrix C is assigned a random integer independently drawn from a uniform distribution of the range [1, 1000]. The triangle inequality $c(i,j) + c(j,k) \geq c(i,k)$ is not assumed in the instances. Although this type of problem instances is application-free, it is mathematically significant. A TSP instance without triangle inequality cannot be approximated within any constant factor. A heuristic local search algorithm usually performs much worse for this type of TSP instances, which offers a strikingly challenge to solving them [2, 3, 6, 10, 11]. We use the 2-opt local search technique in the local search phase. The 2-opt neighborhood can be characterized as the neighborhood that induces the greatest correlation between function values of neighboring tours, because neighboring tours differ in the minimum possible four edges. Along the same reasoning line, the 2-opt may have the smallest expected number of locally optimal points [12]. The local search process randomly selects a solution in the neighborhood of the current solution. A move that gives the first improvement is chosen. The great advantage of the first-improvement pivoting rule is to produce randomized locally optimal points. The software program written for the experiments use several different programming languages and are run in PCs with different versions of Window operating system.

3. The edge matrix *E*

Usually the edge matrix E is not necessary to be included in the TSP definition because the TSP is a complete graph. However, the edge matrix E is an effective data structure that can help us understand the search behavior of a local search system. General local search algorithm may not require much problem-specific knowledge in order to generate good solutions. However, it may be unreasonable to expect a search algorithm to be able to solve any problem without taking into account the data structure and properties of the problem at hand.

To solve a problem, the first step is to create a manipulatable description of the problem itself. For many problems, the choice of data structure for representing a solution plays a critical role in the analysis of search behavior and design of new search algorithm. For the TSP, a tour can be represented by an ordered list of nodes or an edge configuration of a tour in the edge matrix E, as illustrated in **Figure 3**. The improvement of the current tour represents the change in the order of the nodes or the edge configuration of a tour.

Observing the behavior of search trajectories in a local search system can be quite challenging. The edge matrix E is a natural data structure that can help us trace the search trajectories and understand the dynamics of a local search system. An edge $e(i,j)$ is the most basic element of a tour, but contains a piece of information about each of $(n-2)!$ tours that go through it. Essentially, the nature of local search for the TSP is an edge-selection process: preservation of good edges and rejection of bad edges according to the objective function $f(s)$. Each edge has an implicit probability to be selected by a locally optimal tour. A better edge has higher probability to be included in a locally optimal tour. Therefore, the edges in E can be divided into three groups: globally superior edges, G-edges, and bad edges. A globally superior edge is the edge that occurs in many or all locally optimal tours. Although each of these locally optimal tours selects this edge based on its own

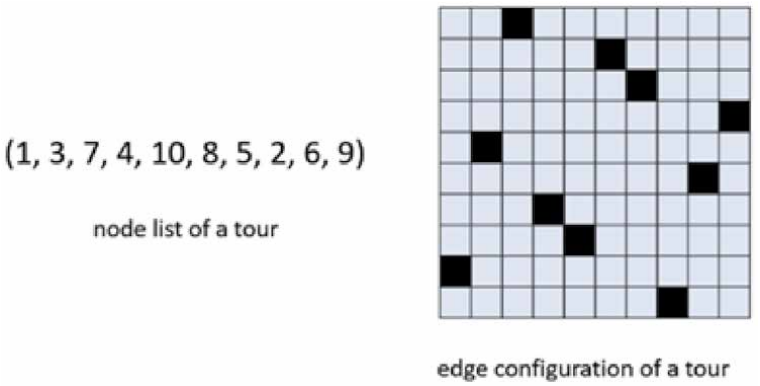


Figure 3.
Two representations of a tour: an ordered list of nodes and an edge configuration of a tour.

search trajectory, the edge is globally superior since the edge is selected by these individual tours from different search trajectories going through different search regions. The globally superior edges have higher probability to be selected by a locally optimal tour. A *G*-edge is the edge that is included in a globally optimal tour. All *G*-edges are globally superior edges and can be treated as a special subset of the globally superior edges. The edges that are discarded by all search trajectories or selected by only few locally optimal tours are bad edges. A bad edge is impossible to be included in a globally optimal tour. A locally optimal tour usually consists of some *G*-edges, some globally superior edges and a few bad edges.

The changes of the edge configuration of the matrix E represent the transformations of the search trajectories in a local search system. When all search trajectories reach their end points, the final edge configuration of E represents the final state of the local search system. For a tour s_k, we define an element $e(i,j)$ of E as

$$e(i,j) = \begin{cases} 1 \text{ if the element } e(i,j) \text{ is in the tour } s_k \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

Then the hit-frequency value e_{ij} in the element $e(i,j)$ is defined as the number of occurrence of the element in K tours, that is

$$e_{ij} = \sum_{k=1}^{K} e(i,j)_k \tag{3}$$

When K search trajectories reach their end points, the value $(e_{ij} + e_{ji})/K$ can represent the probability of the edge $e(i,j)$ being hit by a locally optimal tour. We can use graphical technique to observe the convergent behavior of the search trajectories through the matrix E. The hit-frequency value e_{ij} can be easily converted into a unit of half-tone information in a computer, a value that we interpret as a number H_{ij} somewhere between 0 and 1. The value 1 corresponds to black color, 0 to white color, and any value in between to a gray level. Let K be the number of search trajectories, the half-tone information H_{ij} on a computer screen can be represented by the hit-frequency e_{ij} in the element $e(i,j)$ of E:

$$H_{ij} = \frac{e_{ij}}{K} \tag{4}$$

Figure 4 illustrates a simple example of visualization showing the convergent behavior of 100 search trajectories for a 50-node instance. **Figure 4(a)** shows the image of the edge configurations of 100 random initial tours. Since each element of E has equal chance to be hit by these initial tours, almost all elements are hit by these initial tours, and all elements have very low H_{ij} values, ranging from 0.00 to 0.02. When the local search system starts searching, the search trajectories constantly change their edge configurations, and therefore the colors in the elements of E are changed accordingly. As the search continues, more and more elements become white (i.e. they are discarded by all search trajectories) and other elements become darker (i.e. they are selected by more search trajectories). When all search trajectories reach their end points, the colored elements represent the final edge configuration of the search system. **Figure 4(b)** and **(c)** show the images of edge configuration of E when all search trajectories completed 2000 iterations and 5000 iterations, respectively. At 5000th iteration, the range of H_{ij} values in the elements of E is from 0.00 to 0.42. The value 0.42 means that 42% of the search trajectories select this element. Majority of the elements of E become white color.

This simple example has great explanatory power about the global dynamics of the local search system for the TSP. As search trajectories continue searching, the number of edges hit by them becomes smaller and smaller, and better edges are hit by more and more search trajectories. This edge-convergence phenomenon means that all search trajectories are moving closer and closer to each other, and their edge configurations become increasingly similar. This phenomenon describes the globally asymptotic behavior of the local search system.

It is easily verified that under certain conditons, a local search system is able to find the set of the globally optimal tours S^* when the number of search trajectories is unlimited, i.e.

$$\lim_{K\to\infty} P[S^* \subset S] = 1 \tag{5}$$

However, the required search effort may be very huge – equivalent to enumerating all tours in the solution space. Now one question for the ABSS is "How many search trajectories in the search system do we need to find all globally optimal tours?" The matrix E consists of $n(n-1)$ elements (excluding the diagonal elements). When we randomly construct a tour and record its edge configuration in E, n elements of E will be hit by this tour. If we construct more random tours and record their edge configurations in E, more elements will be hit. We define K as the number of randomly-constructed initial tours, whose edge configurations together

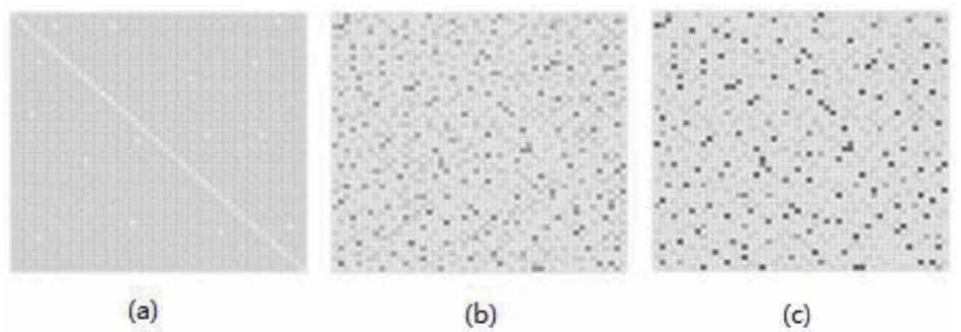


Figure 4.
Visualization of the convergent dynamics of local search system. (a) the image of the edge configurations of 100 initial tours, (b) and (c) the images of edge configurations when the search trajectories are at 2000th and 5000th iteration, respectively.

will hit all elements of E. We know that all elements of E represent all combinatorial possibilities in the solution space. Therefore, K is the number of search trajectories such that the union of edge configurations of ther initial tours covers the entire solution space. In our experiments, we found that the edge configurations of at most $6n$ randomly-constructed tours can guarantee to hit all elements of E. From the tour perspective, $K = 6n$ random tours represent only a small set of the tours in the solution space. However, from the view of edge-configuration, the union of the edge configurations of $6n$ random tours represents the edge configurations of all tours in the solution space. It reveals an amazing fact: the union of the edge configurations of only $6n$ random tours contains the edge configurations of all $n(n-1)!/2$ tours in the solution space. It reflects the combinatorial nature of the TSP: the tours in the solution space are formed by different combinations of the edges. The union of the edge configurations of a set of tours contains information about many other tours because one tour shares its edges with many other tours. One fundamental theory that can help us explain this phenomenon is the information theory [13]. According to the information theory, each solution point contains some information about its neighboring solutions that can be modeled as a function, called *information function* or *influence function*. The influence function of the i^{th} solution point in the solution space S is defined as a function $\Omega_i : S \rightarrow \Re$, such that Ω_i is a decreasing function of the distance from a solution point to the i^{th} solution point. The notion of influence function has been extensively used in datamining, data clustering, and pattern recognition.

4. The nature of heuristic local search

Heuristic local search is based on the concept of neighborhood search. A neighborhood of a solution s_i, denoted as $N(s_i)$, is a set of solutions that are in some sense close to s_i. For the TSP, a neighborhood of a tour s_i is defined as a set of tours that can be reached from s_i in one single transition. From edge-configuration perspective, all tours in $N(s_i)$ are very similar because they share significant number of edges with s_i. The basic operation of local search is iterative improvement, which starts with an initial tour and searches the neighborhood of the current tour for a better tour. If such a tour is found, it replaces the current tour and the search continues until no improvement can be made. The local search algorithm returns a locally optimal tour.

The behavior of a local search trajectory can be understood as a process of iterating a search function $g(s)$. We denote s_0 as an initial point of search and $g^t(s)$ as the t^{th} iteration of the search function $g(s)$. A search trajectory $s_0, g(s_0), g^2(s_0), \ldots, g^t(s_0), \ldots$ converges to a locally optimal point s' as its limit, that is,

$$g\left(\lim_{t\rightarrow\infty} g^t(s_0)\right) = \lim_{t\rightarrow\infty} g^{t+1}(s_0) = s' \tag{6}$$

Therefore, a search trajectory will reach an end point (a locally optimal point) and will stays at this point forever.

In a heuristic local search algorithm, there is a great variety of ways to construct initial tour, choose candidate moves, and define criteria for accepting candidate moves. Most heuristic local search algorithms are based on randomization. In this sense, a heuristic local search algoorithm is a randomized system. There are no two search trajectories that are exactly alike in such a search system. Different search

trajectories explore different regions of the solution space and stop at different final points. Therefore, local optimality depends on the initial points, the neighborhood function, randomness in the search process, and time spent on search process. On the other hand, however, a local search algorithm essentially is deterministic and not random in nature. If we observe the motion of all search trajectories, we will see that the search trajectories go towards the same direction, move closer to each other, and eventually converge into a small region in the solution space.

Heuristic local search algorithms are essentially in the domain of dynamical systems. A heuristic local search algorithm is a discrete dynamical system, which has a solution space S (the state space), a set of times T (search iterations), and a search function $g : S \times T \rightarrow S$ that gives the consequents to a solution $s \in S$ in the form of $s_{t+1} = g(s_t)$. A search trajectory is the sequence of states of a single search process at successive time-steps, which represents the part of the solution space searched by this search trajectory. The questions about the behavior of a local search system over time are actually the questions about its search trajectories. The most basic question about the search trajectories is "Where do they go in the solution space and what do they do when they get there?"

The attractor theory of dynamical systems is a natural paradigm that can be used to describe the search behavior of a heuristic local search system. The theory of dynamical systems is an extremely broad area of study. A dynamical system is a model of describing the temporal evolution of a system in its state space. The goal of dynamical system analysis is to capture the distinctive properties of certain points or regions in the state space of a given dynamical system. The theory of dynamical systems has discovered that many dynamical systems exhibt attracting behavior in the state space [14–22]. In such a system, all initial states tend to evolve towards a single final point or a set of points. The term *attractor* is used to describe this single point or the set of points in the state space. The attractor theory of dynamical systems describes the asymptotic behavior of typical trajectories in the dynamical system. Therefore, the attractor theory provides the theoretical foundation to study the search behavior of a heuristic lcoal search system.

In a local search system for the TSP, no matter where we start a search trajectory in the solution space, all search trajectories will converge to a small region in the solution space for a unimodal TSP instance or h small regions for a h-model TSP. We call this small region a *solution attractor* of the local search system for a given TSP instance, denoted as A. Therefore, the solution attractor of a local search system for the TSP can be defined as an invariant set $A \subset S$ consisting of all locally optimal tours and the globally optimal tours. A single search trajectory typically converges to either one of the points in the solution attractor. A search trajectory that is in the solution attractor will remain within the solution attractor forward in time. Because a globally optimal tour s^* is a special case of locally optimal tours, it is undoubtedly embodied in the solutioin attractor, that is, $s^* \in A$. For a h-modal TSP instance, a local search system will generate h solution attractors $(A_1, A_2, \ldots, A_h)$ that attract all search trajectories. Each of the solution attractors has its own set of locally optimal tours, surrounding a globally optimal tour s_i^* $(i = 1, 2, \ldots, h)$. A particular search trajectory will converge into one of the h solution attractors. All locally optimal tours will be distributed to these solution attractors. According to dynamical systems theory [20], the closure of an arbitrary union of attractors is still an attractor. Therefore, the solution attractor A of a local search system for a h-modal TSP is a complete collection of h solution attractors $A = A_1 \cup A_2 \cup \ldots \cup A_h$.

The concept of solution attractor of local search system describes where the search trajectories actually go and where their final points actually stay in the solution space. **Figure 5** visually summarizes the concepts of search trajectories and solution attractors in a local search system for a multimodal optimization problem,

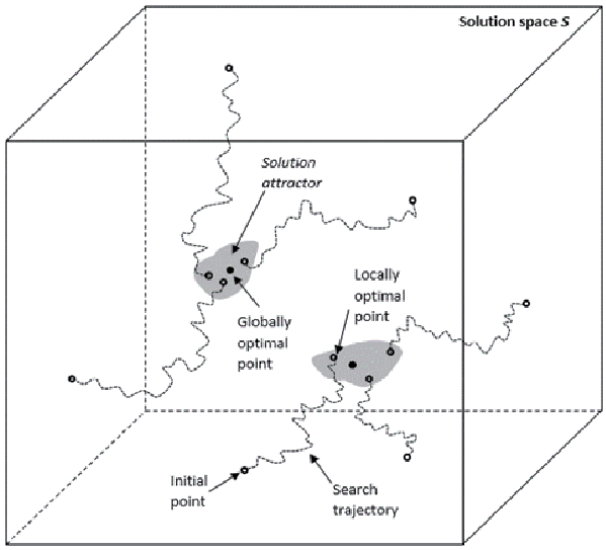


Figure 5.
Illustration of the concepts of serch trajectories and solution attractors in a local search system for a multimodal optimization problem.

describing how search trajectories converge and how solution attractors are formed. In summary, let $g(s)$ be a search function in a local search system for the TSP, the solution attractor of the search system has the following properties [23–25]:

1. *Convexity*, i.e. $\forall s_i \in S, g^t(s_i) \in A$ for sufficient long t;

2. *Centrality*, i.e. the globally optimal tour s_i^* is located centrally with respect to the other locally optimal tours in A_i $(i = 1, 2, \ldots h)$;

3. *Invariance*, i.e. $\forall s' \in A, g^t(s') = s'$ and $g^t(A) = A$ for all time t;

4. *Inreducibility*, i.e. the solution attractor A contains a limit number of invariant locally optimal tours.

A search trajectory in a local search system changes its edge configuration during the search according to the objective function $f(s)$ and its neighborhood structure. The matrix E can follow the "footprints" of search trajectories to capture the dynamics of the local search system. When all search trajectories reach their end points – the locally optimal tours, the edge configuration of the matrix E will become fixed, which is the edge configuration of the solution attractor A. This fixed edge configuration contains two groups of edges: the edges that are not hit by any of locally optimal tours (non-hit edges) and the edges that are hit by at least one of the locally optimal tours (hit edges). **Figure 6** shows the edge grouping in the edge configuration of E when all search trajectories stop at their final points.

In the ABSS, we use K search trajectories in the local search phase. Different sets of K search trajectories will generate different final edge configuration of E. Suppose that, we start the local search from a set of K initial points and obtain a edge configuration M_a in E when the local search phase is terminated. Then we start the local search process again from a different set of K initial points and obtains a little different edge configuration M_b in E. Which edge configuration truly describes the edge configuration of the real solution attractor? Actually, M_a and M_b are structurally

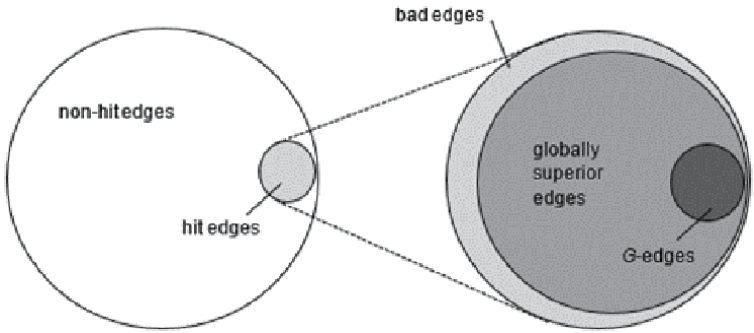


Figure 6.
The grouping of the edges in E *when all search trajectories reach their end points.*

equivalent because they are different only in the set of bad edges, thus M_a precisely replicates the dynamical properties of M_b. The final edge configuration of the constructed solution attractor generated from K search trajectories is not sensitive to the selection of K search trajectories. This property indicates that a heuristic local search system actually is a deterministic system: although a single search trajectory appears stochastic, all search trajectories from differeitn initial points will be always trapped into the same small region in the solution space and the final edge configuration of E will always converge to the same set of the globally optimal edges.

The convergence of the search trajectories can be measured by the change in the edge configuration of the matrix E. In the local search process, search trajectories collect all available topology information about the quality of the edges from their search experience and record such information in the matrix E. The changes in the edge configuration of E fully reflects the real search evolution of the search system. A state of convergence is achieved once no any more local search trajectory can change the edge configuration of E. For a set of search trajectories to be converging, they must be getting closer and closer to each other, that is, their edge configurations become increasingly similar. As a result, the edge configurations of the search trajectories converge to a small set of edges that contains all globally superior edges and some bad edges. Let W denote total number of edges in E, $\alpha(t)$ the number of the edges that are hit by all search trajectories at time t, $\beta(t)$ the number of the edges that are hit by one or some of the search trajectories, and $\gamma(t)$ the number of edges that have no hit at all, then at any time t, we have

$$W = \alpha(t) + \beta(t) + \gamma(t) \tag{7}$$

For a given TSP instance, W is a constant value $W = n(n-1)/2$ for a symmetric instance or $W = n(n-1)$ for an asymmetric instance. During the local search process, the values for $\alpha(t)$ and $\gamma(t)$ will increase and the value for $\beta(t)$ will decrease. However, these values cannot increase or decrease foreover. At certain point of time, they will become constant values, that is,

$$W = \lim_{t\to\infty}\alpha(t) + \lim_{t\to\infty}\beta(t) + \lim_{t\to\infty}\gamma(t) = \mathrm{A} + \mathrm{B} + \Gamma \tag{8}$$

Our experiments confirmed this inference about $\alpha(t)$, $\beta(t)$ and $\gamma(t)$. **Figure 7** illustrates the patterns of $\alpha(t)$, $\beta(t)$ and $\gamma(t)$ curves generated in our experiments. Our experiments also found that, for unimodal TSP instances, the ratio $\gamma(t)/W$ could approach to 0.70 quickly for different sizes of TSP instances. For multimodal TSP instances, this ratio depends on the number of the globally optimal points. However, the set of hit edges is still very small.

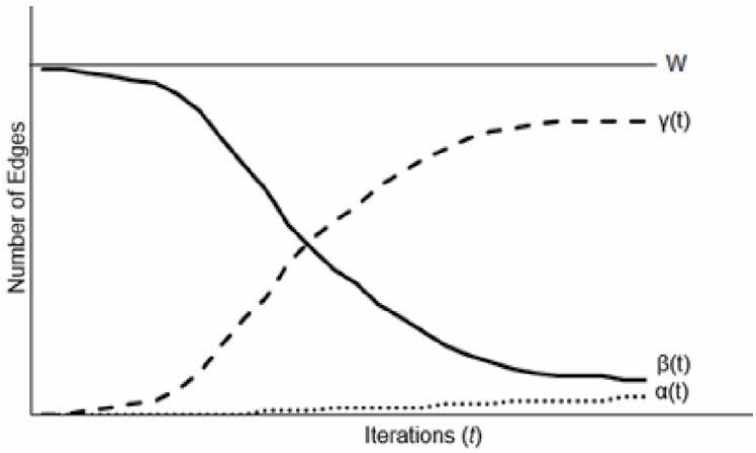


Figure 7.
The $\alpha(t)$, $\beta(t)$ and $\gamma(t)$ curves with search iterations.

In summary, we assume a TSP instance Q has a solution space with h (≥ 1) globally optimal tours $(s_1^*, s_2^*, \ldots, s_h^*)$, and correspondingly there exist h set of G-edges $(G_1, G_2, \ldots, G_h)$. A local search system for the Q will generate h solution attractors $(A_1, A_2, \ldots, A_h)$ that attract all search trajectories. The edge configuration of the solution attractor A is the union of the edge configurations of the h solution attractors. The final edge configuration of E represents the edge configuration of A with three properties:

1. It contains all locally optimal tours;

2. It contains a complete collection of solution attractors, i.e. $A = A_1 \cup A_2 \cup \ldots \cup A_h$;

3. It contains a complete collection of G-edges, i.e. $G = G_1 \cup G_2 \cup \ldots \cup G_h$.

From this analysis, we can see that the edge matrix E is an extremely useful data structure that not only collcets the information about search trajectories, but also convert local search behavor of individual search trajectories into global search behavor of the search system. The global convergence and deterministic property of the search trajectories make the local search system always converge to the same solution attractors and the edge configurations of the search trajectories always converge to the same set of globally superior edges. The matrix E shows us clearly where the search trajectories go and where all locally optimal points are located. We found the village! However, it is still difficult to identify all G-edges among the globally superior edges. The ABSS uses the exhaustive search phase to find all tours in the solution attractor. Since the local search phase has significantly reduced the size of the search space for the exhaustive search phase, the complete search in the solution attractor becomes feasible.

5. Global optimization feature of the ABSS

The task of a global optimization system is to find all absolutely best solutions in the solution space. There are two major tasks performed by a global optimization system: (1) finding all globally optimal points in the solution space and (2) making sure that they are globally optimal. So far we do not have any effective and efficient

global search algorithm to solve NP-hard combinatorial problems. We do not even have well-developed theory or analysis tool to help us design efficient algorithms to perform these two tasks. One critical question in global optimization is how to recognize the globally optimal solutions. Modern search algorithms lack practical criteria that decides when a locally optimal solution is a globally optimal one. What is the necessary and sufficient condition for a feasible point s_i to be globally optimal point? The mathematical condition for the TSP is $\forall s \in S, f(s^*) \leq f(s)$. To meet this condition, an efficient global search system should have the following properties:

1. The search system should be globally convergent.

2. The search system should be deterministic and have a rigorous guarantee for finding all globally optimal solutions without excessive computational burden.

3. The optimality criterion in the system must be based on information on the global behavior of the search system.

The ABSS combines beautifully two crucial aspects in search: exploration and exploitation. In the local search phase, K search trajectories explore the full solution space to identify the globally superior edges, which form the edge configuration of the solution attractor. These K search trajectories are independently and invidually executed, and therefore they create and maintain diversity from beginning to the end. The local search phase is a randomized process due to randomization in the local search function $g(s)$. In this sense, the K search trajectories actually perform the Monte Carlo simulation to sample locally optimal tours. The essential idea of Monte Carlo method is using randomness to solve problems that might be deterministic in principle [26]. In the ABSS, K search trajectories start a sample of initial points from a uniform distribution over the solution space S, and, through the randomized local search process, generate a sample of locally optimal points uniformly distributed in the solution attractor A. The edge configuration of E is actually constructed through this Monte Carlo sampling process.

Each of the K search trajectories passes through many neighborhoods on its way to the final point. For any tour s_i, the size of $N(s_i)$ is greater than $\left(\frac{n}{2}\right)!$ [12]. Let $N(s_i')$ denote the neighborhood of the final point s_i' of the i^{th} search trajectory and $\Omega N(s_{tran})_i$ as the union of the neighborhoods of all transition points of the search trajectory, then we can believe that the search space covered by K search trajectories is

$$N(s_1') \cup \Omega \mathrm{N}(s_{tran})_1 \cup \mathrm{N}(s_2') \cup \Omega \mathrm{N}(s_{tran})_2 \ldots \cup \mathrm{N}(s_K') \cup \Omega \mathrm{N}(s_{tran})_K = S \tag{9}$$

That is, the solution attractor A is formed through the entire solution space S. The solution attractor A contains h unique minimal "convex" sets A_i $(i = 1, 2, \ldots, h)$. Each A_i has a unique best tour s_i^* surrounded by a set of locally optimal tours. The tour s_i^* in A_i satisfies $f(s_i^*) < f(s)$ for all $s \in A_i$ and $f(s_1^*) = f(s_2^*) = \ldots = f(s_h^*)$.

We see that the matrix E plays a critical role to transform local search process of the individual search trajectories into a collective global search process of the system. Each time when a local search trajectory finds a better tour and updates the edge configuraton of E, the conditional distribution on the edges are updated. More values are attached to the globally superior edges, and bad edges are discarded. Let W be the complete set of the edges in E and W_A the set of edges in the edge configuration of the solution attractor A such that $g(W)$ is contained in the interior of W. Then the intersection W_A of the nested sequence of sets is

$$W \supset g(W) \supset g^2(W) \supset \ldots \supset g^t(W) \supset \ldots \supset W_A \tag{10}$$

and $\lim_{t\to\infty} g^t(W_A) = W_A$. As a result, the edge configurations of K search trajectories converge to a small set of edges.

The "convexity" property of the solution attractor A allows the propagation of the minimum property of s_i^* in the solution attractor A_i to the whole solution space S through the following conditions:

1. $\forall s \in A_i, f(s_i^*) < f(s)$

2. $f(s_1^*) = f(s_2^*) = \ldots = f(s_h^*)$

3. $\min_{s\in A} f(s) = \min_{s\in S} f(s)$

Therefore the global convergence and deterministic property of the search trajectories in the local search phase make the ABSS always find the same set of globally optimal tours. We conducted several experiments to confirm this argument empirically. In our experiments, for a given TSP instance, the ABSS performed the same search process on the instance several times, each time using a different set of K search trajectories. The ABSS outputed the same set of the best tours in all trials.

Table 1 shows the results of two experiments. One experiment generated $n = 1000$ instance Q_{1000}, the other generated $n = 10000$ instance Q_{10000}.

Trial number	Number of tours in A	Range of tour cost	Number of best tours in A
Q_{1000} (6000 search trajectories)			
1	6475824	[3241, 4236]	1
2	6509386	[3241, 3986]	1
3	6395678	[3241, 4027]	1
4	6477859	[3241, 4123]	1
5	6456239	[3241, 3980]	1
6	6457298	[3241, 3892]	1
7	6399867	[3241, 4025]	1
8	6423189	[3241, 3924]	1
9	6500086	[3241, 3948]	1
10	6423181	[3241, 3867]	1
Q_{10000} (60000 search trajectories)			
1	8645248	[69718, 87623]	4
2	8657129	[69718, 86453]	4
3	8603242	[69718, 86875]	4
4	8625449	[69718, 87053]	4
5	8621594	[69718, 87129]	4
6	8650429	[69718, 86978]	4
7	8624950	[69718, 86933]	4
8	8679949	[69718, 86984]	4
9	8679824	[69718, 87044]	4
10	8677249	[69718, 87127]	4

Table 1.
Tours in constructed solution attractor A *for* Q_{1000} *and* Q_{10000}.

We conducted 10 trials on each of the instances respectively. In each trial, the ABSS used $K = 6n$ search trajectories. Each search trajectory stopped when no improvement was made during $10n$ iterations. The matrix E stored the edge configurations of the K final tours and then was searched completely using the depth-first tree search process. **Table 1** lists the number of tours found in the constructed solution attractor A, the cost range of these tours, and the number of the best tours found in the constructed solution attractor. For instance, in trial 1 for Q_{1000}, the ABSS found 6475824 tours with the cost range [3241, 4136] in the constructed solution attractor. There was a single best tour in the solution attractor. The ABSS found the same best tour in all 10 trials. For the instance Q_{10000}, the ABSS found the same set of four best tours in all 10 trials. These four best tours have the same cost value, but with different edge configurations. If any trial had generated a different set of the best tours, we could immediately make a conclusion that the best tours in the constructed solution attractor may not be the globally optimal tours. From practical perspective, the fact that the same set of the best tours was detected in all trials provides an empirical evidence of the global optimality of these tours. The fact also indicates that the ABSS converges in solution. *Convergence in solution* means that the search system can identify all optimal solutions repeatedly. Always finding the same set of optimal solutions actually is the fundamental requirement for global optimization systems.

6. Computing complexity of the ABSS

With current search technology, the TSP is an infeasible problem because it is not solvable in a reasonable amount of time. Faster computers will not help. A feasible search algorithm for the TSP is one that comes with a guarantee to find all best tours in time at most proportional to n^k for some power k. The ABSS can guarantee to find all globally optimal tours for the TSP. Now the question is how efficient it is?

The core idea of the ABSS is that, if we have to use exhaustive search to confirm the globally optimal points, we should first find a way to quickly reduce the effective search space for the exhaustive search. When a local search trajectory finds a better tour, we can say that the local search trajectory finds some better edges. It is an inclusive view. We also can say that the local search trajectory discards some bad edges. It is an exclusive view. The ABSS uses the exclusive strategy to conquer the TSP. The local search phase in the ABSS quickly prunes out large number of edges that cannot possibly be included in any of the globally optimal tours. Thus, a large useless area of the solution space is excluded. When the first edge is discarded by all K search trajectories, $(n-2)!$ tours that go through that edge are removed from the search space for the exhaustive search phase. Each time when an edge is removed, large number of tours are removed from the search space. Although the complexity of finding a true locally optimal tour is still open, and we even do not know any nontrivial upper bounds on the number of iterations that may be needed to reach local optimality [27, 28], decades of empirical evidence and practical research have found that heuristic local search converges quickly, within low order polynomial time [1, 8, 27, 29]. In practice, we are rarely able to find perfect locally optimal tour because we simply do not allow the local search process to run enough long time. Usually we let a local search process run a predefined number of iterations, accept whatever tour it generates, and treat it as a locally optimal tour. Therefore, the size of the constructed solution attractor depends not only on the problem structure and the neighborhood function, but also on the amount of search time invested in the local search process. As we increase local search time, we will constructe a smaller and stronger solution attractor. The local search phase in the ABSS can significantly

reduce the search space for the exhaustive search phase by excluding a large number of edges. Usually the local search phase can remove about 60% of edges of the matrix E in $O(n^2)$.

Now an essential question is naturally raised: What is the relationship between the size of the constructed solution attractor and the size of the problem instance? Unfortunately, there is no theoretical analysis tool available in the literature that can be used to answer this question. We have to depend on empirical results to lend some insights. We conducted several experiments to observe the relationship between the size of the constructed solution attractor and the TSP instance size. **Figures 8–10** show the results of one of our experiments. All other similar experiments reveal the same pattern. In this experiment, we generated 10 unimodal TSP instances in the size from 1000 to 10000 nodes with 1000-node increment. For

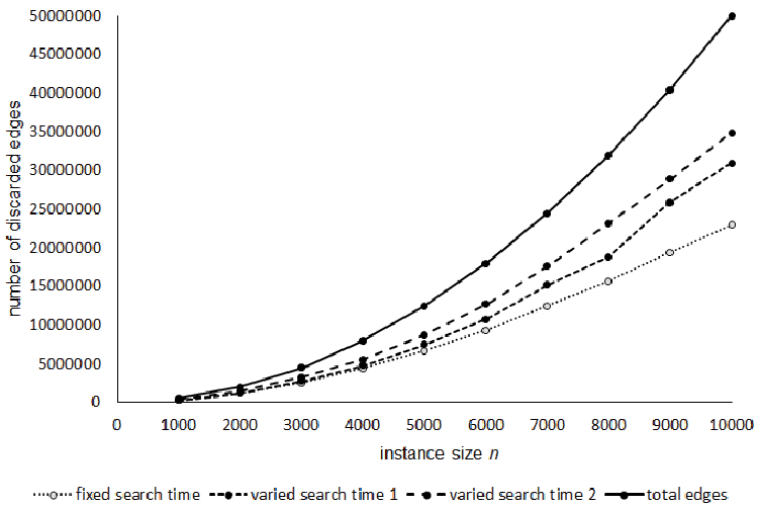


Figure 8.
The number of discarded edges at the end of local search phase.

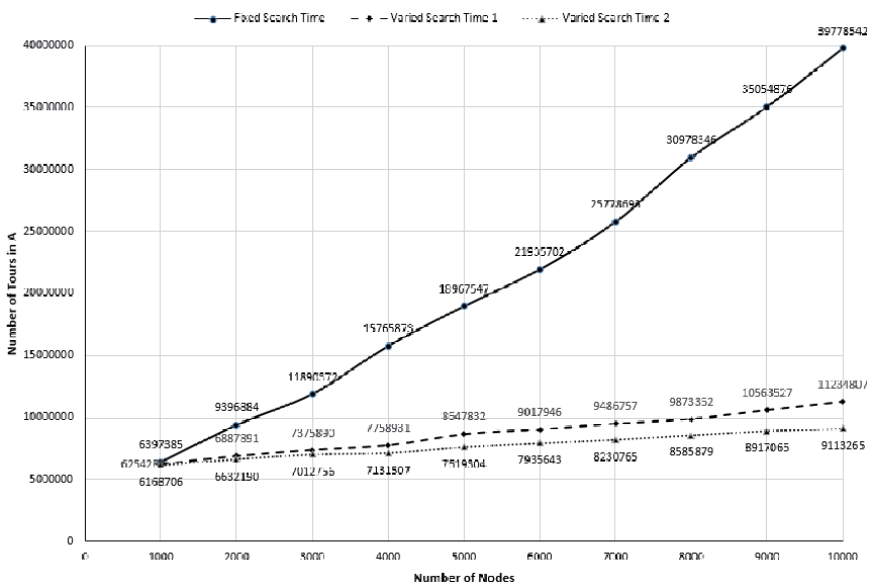


Figure 9.
Relationship between the size of the constructed solution attractor and instance size.

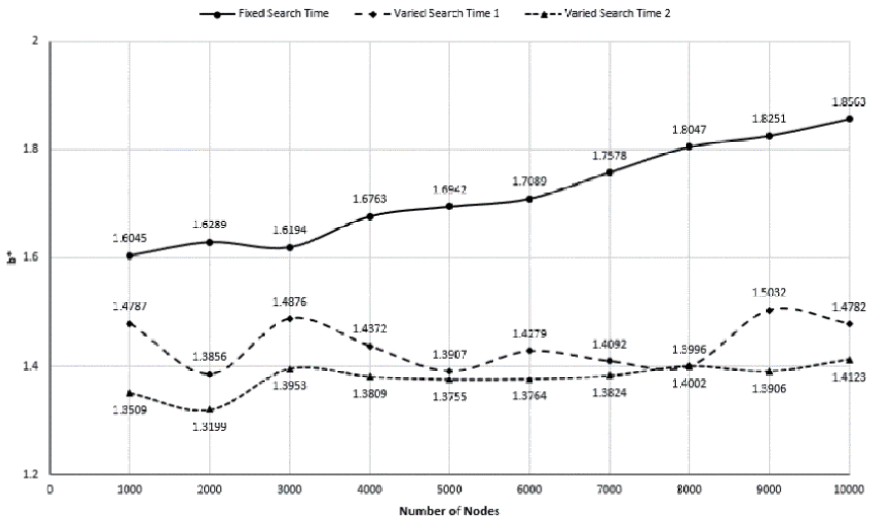


Figure 10.
The b^* values for different instance size n *in our experiment.*

each instance, the ABSS generated $K = 6n$ search trajectories. We first let each search trajectory stop when no tour improvement was made during 10000 iterations regardless of the size of the instance (named "fixed search time"). Then we did the same search procedures on these instances again. This time we made each search trajectory stop when no improvement was made during $10n$ iterations (named "varied search time 1") and $100n$ iterations (named "varied search time 2") respectively. **Figure 8** shows the number of the edges that were discarded at the end of local search phase. **Figure 9** shows the number of tours in the constructed solution attractor for each instance, and **Figure 10** shows the effective branching factors in the exhaustive search phase.

In **Figure 8**, we can see that the search trajectories can quickly converge to a small set of edges. In the fixed-search-time case, about 60% of the edges were discarded by search trajectories for the 1000-node instance, but this percentage decreases as instance size increases. For the 10000-node instance, only about 46% of the edges are discarded. However, if we increase the local search time linearly when the instance size increases, we can keep the same percentage of discarded-edge for all instance sizes. In the varied-search-time-1 case, about 60% of the edges are abandoned for all different instance sizes. In the varied-search-time-2 case, this percentage increases to 68% for all instances. Higher percentage of abandoned edges means that majority of the branches are removed from the search tree.

Figure 9 shows the number of tours exist in the constructed solution attractor for these instances. All curves in the chart appear to be linear relationship between the size of constructed solution attractor and the size of the problem instance, and the varied-search-time curves have much flatter slope because longer local search time makes a smaller constructed solution attractor. **Figures 8** and **9** indicate that the search trajectories in the local search phase can effectively and efficiently reduce the search space for the exhaustive search, and the size of the solution attractor increases linearly as the size of the problem instance increases. Therefore, the local search phase in the ABSS is an efficiently asymptotical search process that produces an extremely small search space for further exhaustive search.

The completely searching of the constructed solution attractor is delegated to the exhaustive search phase. This phase may still need to examine tens or hundreds of millions of tours but nothing a computer processor cannot handle, as opposed to the

huge number of total possibilities in the solution space. The exhaustive search phase can find the exact globally optimal tours for the problem instance after a limited number of search steps.

The exhaustive search phase can use any enumerative technique. However, the edge configuration of E can be easily searched by the depth-first tree search algorithm. One of the advantages of depth-first tree search is less memory requirement since only the nodes on the current path are stored. When using tree-search algorithm, we usually use branching factor, average branching factor, or effective branching factor to measure the computing complexity of the algorithm [30–33]. In the data structure of search tree, the branching factor is the number of successors generated by a given node. If this value is not uniform, an average branching factor can be calculated. An effective branching factor b^* is the number of sucessors generated by a typical node for a given tree-search problem. We use the following definition to calculate effective brancing factor b^* for the exhaustive search phase:

$$N = b^* + (b^*)^2 + \ldots + (b^*)^n \tag{11}$$

where n is the size of the TSP instance, representing the depth of the tree, and N is total number of nodes generated in the tree from the origin node. In our experiments, the tree-search process always starts from node 1 (the first row of E). N is total number of nodes that are processed to construct all valid tours and incomplete (therefore abandoned) tours in E. N does not count the node 1 (the origin node), but includes the node 1 as the end node of a valid tour. We use **Figure 2(d)** as an example. The depth-first search process searches the edge configuration of E and will generate $N = 58$ nodes. Therefore, $b^* \approx 1.3080$, that is, $58 \approx 1.3080 + 1.3080^2 + \ldots + 1.3080^{10}$. **Figure 10** shows the effective branching factor b^* in our experiment. The low values of b^* indicates that the edge configuration of the solution attractor represents a tree with extremely sparse branches, and the degree of sparseness does not changes as the problem size increase if we linearly increase local search time in the local search phase for a large instance. The search time in the exhaustive search phase is probably in $O(n^2)$ since the size of the constructed solution attractor might be linearly increased with the problem size n and the number of edges in E is polynomially increased with the problem size. Our experiments shows that the ABSS can significantly reduce the computational complexity for the TSP and solve the TSP efficiently with global optimality guarantee.

Therefore, the ABSS is a simple algorithm that increases in computational difficulty polynomially with the size of the TSP. In the ABSS, the objective pursued by the local search phase is "quickly eliminating unnecessary search space as much as possible." It can provide an answer to the question "In which small region of the solution space is the optimal solution located?" in time of $O(n^2)$. The objective of the exhaustive search phase is "identifying the best tour in the remaining search space." It can provide an anwer to the question "Which is the best tour in this small region?" in time of $O(n^2)$. All together, the ABSS can answer the question "Is this tour the best tour in the solution space?" in time of $O(n^2)$. Therefore, the ABSS is probably with computing complexity of $O(n^2)$ and memory space requirement of $O(n^2)$. This suggests that the TSP might not be as complex as we might have expected.

7. Conclusion

Advances in computational techniques on the determination of the global optimum for an optimization problem can have great impact on many scientific and

engineering fields. Although both the TSP and heuristic local search algorithms have huge literature, there is still a variety of open problems. Numerous experts have made huge advance on the TSP research, but two fundamental questions of the TSP remain essentially open: "How can we find the optimal tours in the solution space, and how do we know they are optimal?"

The P-vs-NP problem is about how fast we can search through a huge number of solutions in the solution space [34]. Do we ever need to explore all the possibilities of the problem to find the optimal one? Actually, the P-vs-NP problem asks whether, in general, we can find a method that completely searches only the region where the optimal points are located [34–36]. Most people believe P $\neq$ NP because we have made little fundamental progress in the area of exhaustive search. Modern computers can greatly speed up the search, but the extremely large solution space would still require geologic search time to find the exact optimal solution on the fastest machines imaginable. A new point of view is needed to improve our capacity to tackle these difficulty problems. This paper describe a new idea: using efficient local search process to effectively reduce the search space for exhaustive search. The concept of solution attractor in heuristic local search systems may change the way we think about both local search and exhaustive search. Heuristic local search is an efficient search system, while exhaustive search is an effective search system. The key is how we combines these two systems into one system beautifully to conquer the fundamental issues of the hard optimization problems. In the TSP case, the edge matrix E, a problem-specific data structure, plays a critical role of reducing the search space and transforming local search to global search.

The ABSS is designed for the TSP. However, the concepts and formulation behind the search algorithm can be used for any combinatorial optimization problem requiring the search of a node permutation in a graph.

Author details

Weiqi Li
School of Management, University of Michigan-Flint, Flint, USA

*Address all correspondence to: weli@umich.edu

References

[1] Applegate DL, Bixby RE, Chaátal V, Cook WJ. The Traveling Salesman Problem: A Computational Study. Princeton: Princeton University Press; 2006

[2] Papadimitriou CH, Steiglitz K. Combinatorial Optimization: Algorithms and Complexity. New York: Dover Publications; 1998

[3] Papadimitriou CH, Steiglitz K. On the complexity of local search for the traveling salesman problem. SIAM Journal on Computing. 1977:6:76–83

[4] Gomey J. Stochastic global optimization algorithms: a systematic formal approach. Information Science. 2019:472:53–76

[5] Korte B, Vygen J. Combinatorial Optimization: Theory and Algorithms. New York: Springer; 2007

[6] Rego C, Gamboa D, Glover F, Osterman C. Traveling salesman problem heuristics: leading methods, implementations and latest advances. European Journal of Operational Research. 2011:211:427–411

[7] Zhigliavsky A, Zillinakas A. Stochastic Global Optimization. New York: Springer; 2008

[8] Aart E, Lenstra JK. Local Search in Combinatorial Optimization. Princeton: Princeton University Press; 2003

[9] Michalewicz Z, Fogel DB. How to Solve It: Modern Heuristics. Berlin: Springer; 2002

[10] Sahni S, Gonzales T. P-complete approximation problem. Journal of the ACM. 1976:23:555–565

[11] Sourlas N. Statistical mechanics and the traveling salesman problem. Europhysics Letters. 1986:2:919–923

[12] Savage SL. Some theoretical implications of local optimality. Mathematical Programming. 1976:10: 354–366

[13] Shammon CE. A mathematical theory of communication. Bell System Technical Journal. 1948:27:379–423& 623–656

[14] Alligood KT, Sauer TD, York JA. Chaos: Introduction to Dynamical System. New York: Springer; 1997

[15] Auslander J, Bhatia NP, Seibert P. Attractors in dynamical systems. NASA Technical Report NASA-CR-59858; 1964

[16] Brin M, Stuck G. Introduction to Dynamical Systems. Cambridge: Cambridge University Press

[17] Brown R. A Modern Introduction to Dynamical Systems. New York: Oxford University Press.

[18] Denes A, Makey G. Attractors and basis of dynamical systems. Electronic Journal of Qualitative Theory of Differential Equations. 2011:20(20):1–11

[19] Fogedby H. On the phase space approach to complexity. Journal of Statistical Physics. 1992:69:411–425

[20] Milnor J. On the concept of attractor. Communications in Mathematical Physics. 1985:99(2):177–195

[21] Milnor J. Collected Papers of John Milnor VI: Dynamical Systems (1953–2000). American Mathematical Society; 2010

[22] Ruelle D. Small random perturbations of dynamical systems and the definition of attractor. Communications in Mathematical Physics. 1981:82:137–151

[23] Li W. Dynamics of local search trajectory in traveling salesman problem. Journal of Heuristics. 2005:11: 507–524

[24] Li W, Feng M. Solution attractor of local search in traveling salesman problem: concepts, construction and application. International Journal of Metaheuristics. 2013:2(3): 201–233

[25] Li W, Li X. Solution attractor of local search in traveling salesman problem: computational study. International Journal of Metaheuristics. 2019:7(2):93–126

[26] Kroese DP, Taimre T, Botev ZI. Handbook of Monte Carlo Methods. New York: John Wiley & Sons; 2011

[27] Fischer ST. A note on the complexity of local search problems. Information Processing Letters. 1995:53 (2):69–75

[28] Chandra B, Karloff HJ, Tovey CA. New results on the old-opt algorithm for the traveling salesman problem. SIAM Journal on Computing. 1999:28(6): 1998–2029

[29] Grover, LK. Local search and the local structure of NP-complete problems. Operations Research Letters. 1992:12(4):235–243

[30] Baudet GM. On the branching factor of the alpha-beta pruning algorithm. Artificial Intelligence. 1978: 10(23):173–199

[31] Edelkamp S, Korf RE. The branching factor of regular search space. Proceedings of the 15th National Conference on Artificial Intelligence. 1998:292–304

[32] Korf RE. Depth-first iterative deepening: an optimal admissible tree search. Artificial Intelligence. 1985:27: 97–109

[33] Pearl J. The solution for the branching factor of the alpha-beta pruning algorithm and its optimality. Communication of the ACM. 1982:25 (8):559–564

[34] Fortnow L. The Golden Ticket – P, NP, and the Search for the Impossible. Princeton: Princeton University Press; 2013

[35] Fortnow L. The status of the P versus NP problem. Communication of the ACM. 2009:52(9):78–86

[36] Sipser M. The history of status of the P verses NP question. Proceedings of 24th Annual ACM Symposium on Theory of Computing. 1992:603–618

Chapter 2

Opinion Dynamics and the Inevitability of a Polarised and Homophilic Society

Rafael Prieto Curiel

Abstract

A polarised society is frequently observed among ideological extremes, despite individual and collective efforts to reach a consensual opinion. Human factors, such as the tendency to interact with similar people and the reinforcement of such homophilic interactions or the selective exposure and assimilation to distinct views are some of the mechanisms why opinions might evolve into a more divergent distribution. A complex model in which individuals are exposed to alternating waves of propaganda which fully support different extreme views is considered here within an opinion dynamics model. People exposed to different extreme narratives adopt and share them with their peers based on the persuasiveness of the propaganda and are mixed with their previous opinions based on the volatility of opinions to form a new individual view. Social networks help capture elements such as homophily, whilst persuasiveness and memory capture bias assimilation and the exposure to ideas inside and outside echo chambers. The social levels of homophily and polarisation after iterations of people being exposed to extreme narratives define distinct trajectories of society becoming more or less homophilic and reaching extremism or consensus. There is extreme sensitivity to the parameters so that a small perturbation to the persuasiveness or the memory of a network in which consensus is reached could lead to the polarisation of opinions, but there is also unpredictability of the system since even under the same starting point, a society could follow substantially different trajectories and end with a consensual opinion or with extreme polarising views.

Keywords: Opinion dynamics, polarisation, homophily, consensus, diffusion, interaction network

1. Introduction

Modelling some aspects of our society is challenging at an individual and at a collective level. Every idea, every human feeling and every interaction is so unique that measuring and modelling human constructs such as freedom, love, traditions, friendship, power, or fear is defying from its basis. Obtaining a generalisation or an abstraction, such as physical laws, which apply at a social level is frequently not feasible. Two equal drops of water will act the same under similar circumstances, but no two individuals are so similar as to ensure they feel the same, think the same or react the same to some circumstances. Social settings, as opposed to physical

observed ones, often lack of measuring instruments and units, it is almost impossible to repeat experiments and so transforming our knowledge about society into simple, absolute, and universal descriptions is often unimaginable [1]. Social models are inevitably incomplete and inaccurate, because of scientific limitations and a lack of data [2] and because conventional scientific approaches cannot be applied to many of the problems faced by our society [3]. Furthermore, just a few years ago it was impossible to use the right amount of data or to model more than just a few aspects of the individuals, but today we are capable of simulating large human systems [4] with more complex interactions between its members and its environment [5]; to understand the emergence of crowd behaviour in different situations and to challenge and, in some cases, to measure, some of the theories which are frequently applied across some scientific fields [6]. Models of collective human behaviour have gained interest as the need for them grows, their results get more and more applied in policy and decision-making and their implications are spread throughout more widely.

Models of social behaviour are complex. Many features observed at a social level are an emergent behaviour that results from interactions at a personal level and feedbacks between society and its individuals. Social behaviours are the result of collective individual actions. People adapt rapidly to new circumstances, transforming society as a whole on that process, for instance, by making it normal to maintain some physical distance with others or by wearing a facemask during the COVID-19 pandemic, but some of these social features synchronise our behaviours as well, by the constant feedback others provide.

Modelling society usually requires a substantial level of simplification at the microscopic, individual level in the hope to resemble the macroscopic, social behaviour [7]. The mathematical approach is usually to study the emergent collective patterns when thousands or millions of people -or events- are considered. For instance, a crime might be regarded as a point on a map, a friendship could be considered as a link in a network, or a driver could be modelled by its position and its speed; however, these simplifications made within a social context have helped us to understand the emergent patterns of criminal hotspots [8], the small-world phenomena observed in many social networks [9] or the formation of traffic jams despite efforts from drivers to avoid them [10].

Opinions and the ways they are updated is a complex social system. In general, individuals have an opinion about a specific topic, which is somehow updated when they are confronted with other ideas. Usually, a person gains some confidence in their views when they are reinforced by exposure to similar ideas or challenges their beliefs when they are exposed to different opinions. The exposure to distinct views is a social process and therefore, updating beliefs is mostly a social process as well, which happens perhaps during a simple conversation with others, when listening to what others say on the news, or what they publish on social media. And, as with other complex social systems, individuals transform their society with their opinion, but society transforms individuals as well. There are feedbacks between individual opinions and their collective perceptions and ideas.

1.1 Polarised opinions

Polarisation and the way it emerges is one of the key questions in opinion dynamics models [11]. An increasingly polarised society is observed in attitudes towards the COVID-19 pandemic, views in favour or against a vaccine [12], the consumption of media outlets, opinions on social media and many more. Increased exposure to ideas within an homogeneous community intensifies their tendency to be credulous, whether it is to scientific evidence, unsubstantiated rumours,

inconclusive evidence or even fake news. Polarised opinions might foster confirmation bias, so that people with more extreme opinions tend to become more certain in their beliefs [13] and therefore, it contributes to the proliferation of fake news, whereby once an idea is adopted, is rarely corrected [14].

Frequently, individuals want to persuade others -even unintentionally- to adopt an idea and so there are active efforts to reach a consensual opinion. observing opinion dynamics only at a global scale and ignoring individual dynamics often lead opinions to a consensus state [15], in a similar way in which temperature differences tend to vanish. Yet, two or more contrasting ideas might be highly popular, even if all individual efforts try to reach a consensual opinion. Polarisation, or even fragmentation among many opinions, might be one of the emergent states of collective opinion dynamics, where contrasting ideas might co-exist as a steady state in a society.

Human factors such as the frequency at which we form ties with similar people (homophily), the tendency of having similar opinions as a result of social interactions (social influence), the fact that when presented with mixed evidence, individuals might perceive it as positive feedback for their initial position (biased assimilation), or interpreting the acceptance of an idea as reinforcement when sharing an opinion in a social environment (social feedback) are some of the causal mechanisms why the process in which ideas are updated might be polarising, meaning that final opinions are more divergent than initial opinions [11, 16].

1.2 Homophilic opinions

Usually, a person interacts with others of similar age, income or other sociodemographic, behavioural, and intrapersonal characteristics, including opinions or views on a certain topic [17, 18]. If a population has polarised opinions, it means that, at a global level, there is a high probability that when two individuals are randomly picked, they share extreme different views. However, little is known with respect to the actual interactions. Individuals from a highly polarised society could almost always interact with people who share similar views if polarised bubbles rarely interact with each other. Yet, a different opinion process within the same polarised society is observed when people frequently interact with others who shared opposite views. On a polarised population, opinions have high *homophily* if most of the individuals interact with people with similar views, and opinions are *not homophilic* if people with different views interact frequently. See **Figure 1**, where opinions are represented by the intensity of the colour of a node.

Polarisation between two opinions -or fragmentation among many- is detected when opinions are observed at a global scale, but to detect if opinions are homophilic, more local information with respect to the interactions is needed. For instance, in the 2016 UK referendum to remain in the European Union, 52% of the votes were to leave (a highly polarised election), but at a more local level, the area which voted most heavily in favour of one of the options was Gibraltar (where nearly 96% of the votes were to remain), whereas in Watford results were evenly distributed among leave and remain. Thus, Gibraltar had the lowest polarisation, where there was a near consensus for one of the options, but Watford had the highest polarisation between the leave and remain options. In Watford, however, with their highly polarised election outcome, interactions could still happen very frequently between people with similar views, if the opinion sharing process is highly homophilic and there are little interactions between the two voter groups.

A slightly polarised society does not have homophilic opinions, but a polarised society might have homophilic opinions, or not, depending on how individuals interact and the opinion profile. The relevance of opinion homophily stems from

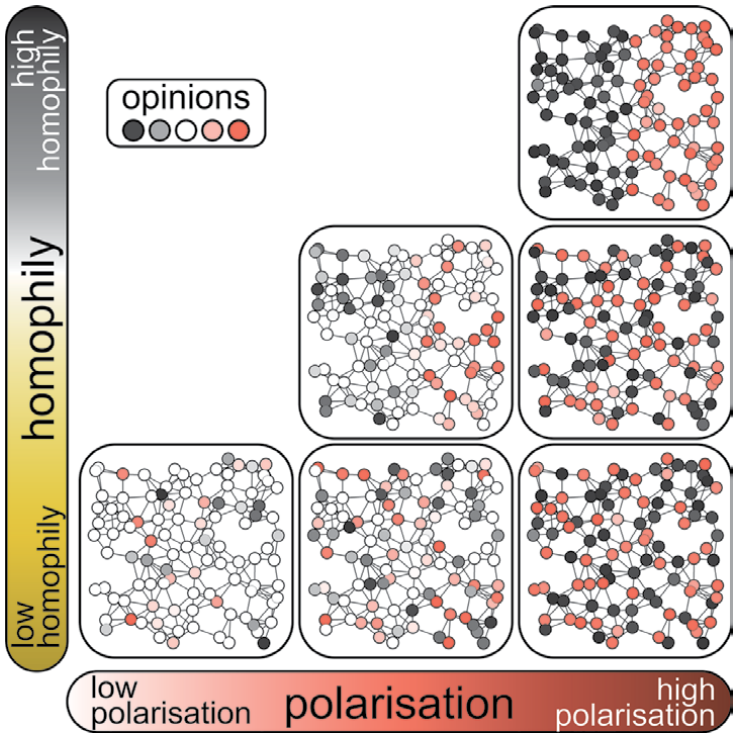


Figure 1.
Opinions (represented by the different colours of the nodes) are shared between individuals who interact (if there is a link between the nodes). Different states in which opinions are distributed show a small polarisation (left part, where most individuals have similar views) or high polarisation (right part, where opinions are split in half) and might show low homophily (bottom part, where opposite opinions are frequently shared among interacting individuals) or high homophily (top part, where opposite opinions are rarely shared among connected nodes).

the fact that in a highly polarised population, most individuals might not be aware that so many people with different views even exist, whereas in a polarised society with little levels of homophily, encounters between people with opposite views happen frequently. Furthermore, a highly polarised society might be a steady state of some opinion dynamics but given the right circumstances (parameters) that state could be highly homophilic or a state in which most individuals interact frequently with people with different views.

Social media and other technological changes could increase exposure to diverse perspectives [19], but at the same time facilitate some mechanisms, such as the creation of links or friendships in the network, filter algorithms and rank information which may accelerate the formation of homophilic communities [16, 20]. People frequently aggregate in groups of interest, and those existent communities frequently adopt narratives from different topics, reinforcing polarisation across distinct themes, for instance, political ideology and perceptions with respect to the COVID-19 pandemic [21, 22]. People interacting with homogeneous communities tend to grow more extreme opinions and become more certain in their beliefs [13] which can favour the spread of misinformation from partisan media and increase animosity within the population [23]. For COVID-19, for example, most of the misinformation detected involves reconfigurations, where existing (often true) facts are adjusted to fit different narratives [24] which are then reproduced by large homophilic groups as facts. Massive misinformation is becoming one of the main threats to our society [14, 25, 26] which might be fostered by an increasingly homophilic opinion dynamic process and a polarised society.

2. Modelling opinions and its dynamics

Opinion formation has been studied from many angles and different mathematical techniques, including mean-field theory and kinetic models of opinion formation [27], or by agents on a social network. Individual opinions on a certain topic are usually modelled as a single-valued number contained in some closed interval which represents extreme (opposing) opinions, for example, left–right leaning voters [28], the level of production of an employee in a plant [7] or perceptions between security and insecurity [29]. The process of opinion updating then is modelled as the result of interaction with other views, a process of self-thinking, some memory loss, or external factors. Interactions between individuals are usually modelled on some social structure, such as a network, considering some spatial proximity, or considering some social aspects, such as the level of influence of one individual to others [30]. A long-term, steady distribution of opinions is usually obtained, either as an analytical solution to some differential equations or through simulations, which reveals among others, the formation of opinion clusters, political segregation [31], vaccine hesitancy [12], the use of certain tools [32], the spread of fear of crime more as a result of opinion dynamics than crime itself [29] or even the diffusion of fake news [14].

2.1 The key ingredients in opinion dynamics models

There are four ingredients in opinion dynamics models [30, 31]:

1. **Individual opinions** - Modelled usually as a number, say $s_i(t) \in [-1, 1]$ for individual i at time t, based on the extremes of an interval, -1 and $+1$, which are identified as opposing opinions, for instance, the levels of support or opposition for an idea, perceptions of security or insecurity, or left–right leaning political views. Other approaches include multi-dimensional views or discrete opinions.

2. **External or individual forces** - External forces such as exposure to news sources [33], or events such as suffering a crime [29] or an accident, and individual forces, such as memory loss may cause an individual to update their views about certain topics.

3. **Updating process as interactions with others** - Exposure to different ideas is frequently considered as the updating mechanism of opinions. Frequently, through interactions with others, person i finds a distinct opinion, $s_j(t)$, and might update their own views according to some function based on their opinion $s_i(t)$ and the "new" one, as $s_i(t+1) = f\big(s_i(t), s_j(t)\big)$, where usually the function f is assumed to get opinion s_i closer to s_j as a result of some consensus effort. Interactions are frequently modelled on some network, where two connected nodes (individuals) share opinions with (some) of its adjacent nodes. The network structure and whether it is directional thus play a role in the updating process.

4. **Metrics** - From the collective opinions, or "opinion profile", $S(t) = (s_1(t), s_2(t), \ldots, s_N(t))$, usually its mean $\overline{S}(t)$ and other metrics are frequently analysed, perhaps dividing by some population groups or node attributes, usually as a long-run behaviour of the dynamics.

Although some analytical results are available [27, 28], the dynamics are usually simulated on a network. The technique allows considering individual aspects, such as assertiveness, persuasiveness, supportiveness, extremists or opinion volatility [28, 31, 34, 35].

2.1.1 Measuring polarisation and homophily

A group might reach an agreement or *consensus* on some opinions if the majority of the individuals share similar views, whereas a group might be *polarised* if opinions are divergent, with *extremism* being the state in which opinions are mostly concentrated among the two extremes. One way to measure the level of polarisation in a population by the variance of the opinion profile, where a large variance means a more polarised society and a small variance means consensus. Formally, the polarisation Φ of an opinion profile S is given by

$$\Phi(S) = Var(S) = \frac{1}{N}\sum_{i=1}^{N}(s_i - \bar{s})^2. \tag{1}$$

For opinions bounded inside the $[-1, +1]$ interval, very small populations could have $\Phi(S)$ values larger than 1, but for a population with more than 100 individuals, $\Phi < 1.01$ and so for large enough populations, it might be considered that Φ obtains values in the 0 (if there is consensus) to 1 (if there is extremism) interval. If a random opinion is sampled for each individual $s_k(0) \in [-1, +1]$, a 95% interval of the polarisation is $\Phi(S) \in [0.327, 0.338]$ and therefore, the distribution of opinions S is classified as *consensus* if $\Phi(S) \approx 0$; *consensual* if $\Phi(S) < 1/3$; *homogeneous* if $\Phi(S) \approx 1/3$, so that the polarisation is similar to a random distribution of opinions; *polarising* if $\Phi(S) > 1/3$ and as *extremism* if $\Phi(S) \approx 1$ (**Figure 2**).

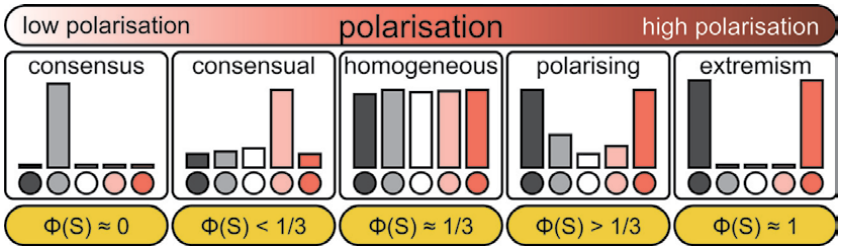


Figure 2.
Classification of collective opinions according to their distribution (represented as the height of each colour bar), from consensus (left) where $\Phi(S) \approx 0$, to extremism (right) where $\Phi(S) \approx 1$.

The process of opinion dynamics has a high level of *homophily* if most of the interactions happen between individuals of similar views and has a low level otherwise. Formally, if $\mathcal{A}_i$ are the adjacent nodes of i, then the average opinion distance $\mathcal{D}_i$ experienced by i is given by

$$\mathcal{D}_i = \frac{1}{d_i}\sum_{j \in \mathcal{A}_i}|s_i - s_j|, \tag{2}$$

where d_i is the degree of i, so that $\mathcal{D}_i$ gives the average opinion distance from a node to its adjacent neighbours (and define $\mathcal{D}_i = 0$ if i has no neighbours). The opinion homophily $\Lambda(S)$ is defined as

$$\Lambda(S) = 1 - \frac{1}{N}\sum_{i=1}^{N} \mathcal{D}_i, \tag{3}$$

a metric suited for measuring homophily based on a continuous node attribute, such as opinions, with high values if individuals interact with others of similar views and has lower values (possibly negative) if interactions are more frequent between individuals of very different views. Notice that the metric depends on the opinion profile but also on the network topology. On a linear network, for instance, where all nodes have two neighbours, except for the two extremes, opinions in the $[-1, +1]$ interval are highly polarised (or have extremism) if half of the individuals have -1 and the other half have $+1$ as opinions, and $\Phi(S) \approx 1$. Such opinion profile is not homophilic with alternating opinions, $S = (+1, -1, +1, -1, \ldots, -1)$, and $\Lambda(S) = -1$, but a high level of homophily is observed when opposite opinions are located on the two extremes of the network, so $S = (+1, +1, \ldots, +1, -1, \ldots, -1, -1)$, in which case, only the two neighbouring individuals located at the boundary between the opinion groups have an interaction with a person who has a distinct opinion than their own and so $\Lambda(S) = 1 - 2/N \approx 1$. The expected opinion distance between two randomly selected opinions is $2/3$, from which $\Lambda(S) \approx 1$ means *preferential interactions* between individuals of similar views; $\Lambda(S) > 1/3$ means *homophilic interactions*; $\Lambda(S) \approx 1/3$ means random interactions; and $\Lambda(S) < 1/3$ means *discouraged interactions* between people of similar views.

2.2 An opinion dynamics model

Consider a diffusion process of opinions on a network, where the four key ingredients (individual opinions, updating process due to individual or external forces, interactions, and the corresponding metrics) are defined as follows. Initially, N individuals have a randomly-distributed opinion $s_i(0) \in [-1, +1]$, which represent two extreme views on a certain topic. As external forces in the dynamics, we consider exposure from the individuals to some "propaganda" in favour of one of the views. Each time step, a randomly selected group of 1% of the individuals are exposed, in an alternative sequence, to some supporting mechanism in favour of any of the two views. It is assumed that the views fully favour one of the two extreme opinions, so that they are considered as $v_1 = +1$, the first force which favours view $+1$, and then $v_2 = -1$, which supports view -1 and so on, with $v_k = -(-1)^k$. As opinion dynamics, individuals who are exposed to any of the views ($v_k = \pm 1$) decide whether to "trust" or to "dismiss" the views based on their current opinion and on the persuasiveness of the views θ, where $\theta \in [0, 1]$ is a parameter which captures how seductive are the views (where large values of θ mean that views are more seductive and individuals are more inclined to trust them and smaller values mean that views are likely to be dismissed). Due to confirmation bias, individuals with opinion closer to $+1$ are more likely to trust a $v_k = +1$ propaganda, as it confirms their views and more likely to ignore $v_l = -1$ propaganda for the same reason. To capture confirmation bias, it is assumed that person i with opinion s_i trusts view v_k with probability

$$\frac{(1 + s_i)\theta}{2} \quad \text{if} \quad v_k = +1, \text{and} \tag{4}$$

$$\frac{(1 - s_i)\theta}{2} \quad \text{if} \quad v_k = -1. \tag{5}$$

With this condition, a person with opinion $s_i = 0.4$ trusts view $v_k = +1$ with probability 0.7θ, but trust view $v_j = -1$ with probability 0.3θ, for some θ which depends on how seductive the corresponding propaganda is, so that individuals are more inclined to trust views which favour their own opinions. Individuals who are seduced by any propaganda, share it with all their contacts as an active effort to persuade them, say by sharing or posting the views on social media. Individuals who dismiss propaganda, make a permanent decision to ignore it, do not update their views and do not share it with their contacts. Thus, when individuals are exposed for the first time to the views, they make a permanent choice whether to accept it (and update their views and share it) or to ignore it (and do nothing). Therefore, after 1% of the individuals are first exposed to propaganda, some individuals trust and share it with their contacts and so on until no one is exposed for the first time to that propaganda. It is assumed that the sharing mechanism (social media, say) works faster than the creation of new propaganda, so that by the time a new view v_{k+1} is created and distributed, the dynamics of the previous (opposing) propaganda v_k has finished. Each wave of propaganda follows a similar diffusion process as the SIR model used in epidemics, where a small percentage of the individuals are initially exposed. The "infection" (the propaganda) passes through individuals, and the distribution of the recovered individuals is observed [12, 32].

Individuals who accept some propaganda at time t update their views according to the *volatility* of their opinions, $\mu \in [0, 1]$, such that individuals who accepted view v_k update their opinion between t and $t + 1$ according to

$$s_i(t + 1) = \mu v_k + (1 - \mu)s_i(t), \tag{6}$$

so that if opinions are volatile (that is, individuals easily update their views, with a large value of μ), then most of their opinion at time $t + 1$ depends only on the views of the propaganda they accepted, but with more rigid opinions (individuals change little their past views, with small μ), then the impact of propaganda becomes small. For example, for view $v_k = +1$ and with volatility $\mu = 0.5$, a person with opinion $s_i(t) = 0.8$ updates their view to $s_i(t + 1) = 0.9$ if they accept v_k, whereas a person with view $s_j(t) = -0.8$ updates their view to $s_j(t + 1) = 0.1$, meaning that a person with very different views from certain propaganda is more difficult to convince, but once convinced, the impact in their opinion is larger (**Figure 3**).

A crucial element in the opinion models is the way in which interactions between individuals are structured. Society has opinion clusters -for example, a

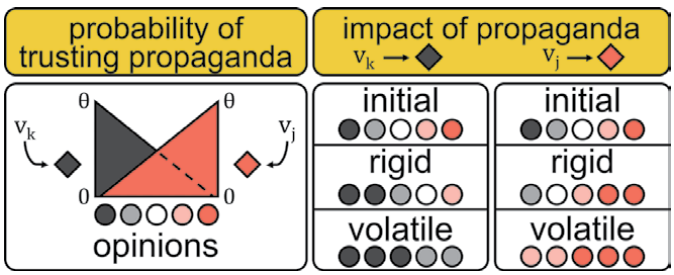


Figure 3.
Probability of trusting any of the two types of propaganda, $v_k = \pm 1$, represented as the two triangles on the left, based on the individual opinions, represented as the colour of the nodes and based on how seductive are two views, θ. Propaganda which supports the views of a person is more likely to be trusted by the person, but still, all propaganda has a certain level of persuasiveness, θ (the maximum height of the triangles). The impact of trusting some propaganda on individual opinions is higher if opinions are more volatile, that is, higher values of μ and has little impact if opinions are more rigid, which is shown as a slight colour change for rigid opinions and a drastic colour change for volatile opinions on the right.

social media group in which information flows easily-, has opinion hubs - influencers, for example, who reach a large population- is likely to be strongly connected with many shortcuts between people who are not directly connected and therefore, the network in which propaganda is shared is also a key element in the model. Four network topologies with $N = 2,000$ nodes are analysed here: (1) a fully connected network; (2) a proximity network (nodes are located randomly on a square and pairs at a distance smaller than a certain threshold are connected); (3) a small-world network and (4) a scale-free network.

The model has two parameters: the persuasiveness θ, which is assumed to be the same for all propaganda, and the volatility of opinions μ, which is assumed to be the same for all individuals. For some values of θ and μ and for some randomly assigned initial opinions, individuals are exposed to a total of 128 waves of propaganda (64 supporting each view).

3. Results

The trajectory of a society in terms of its polarisation Φ and its homophily Λ after each round of propaganda shows that for different network topologies, opinion dynamics yields different states. On a fully connected network, in which there is no relevant network structure, each round of propaganda reaches all individuals (if at least one person trusted it) and seduce some of them based only on their current opinion (**Figure 4**). After some propaganda rounds, most individuals have an opinion either close to $+1$ or to -1 so that polarisation is eventually maximum. Also, since all individuals interact with others, the homophily is reduced when polarisation increases. However, on some other topologies, there is a different impact of each wave of propaganda, particularly after repetition. On a proximity network, most rounds of propaganda tend to increase the level of polarisation, but after repetition, most of the propaganda rounds also increase the level of homophily. Thus, after many rounds, the network has regions with similar (extreme) views and therefore, at a local level, nodes are mostly connected to others with similar views. On a small-world network and a scale-free network, most rounds of propaganda increase the level of polarisation, but the presence of network shortcuts and hubs reduce considerably the homophily so that most of the trajectories are less homophilic than its initial levels after 128 rounds of propaganda.

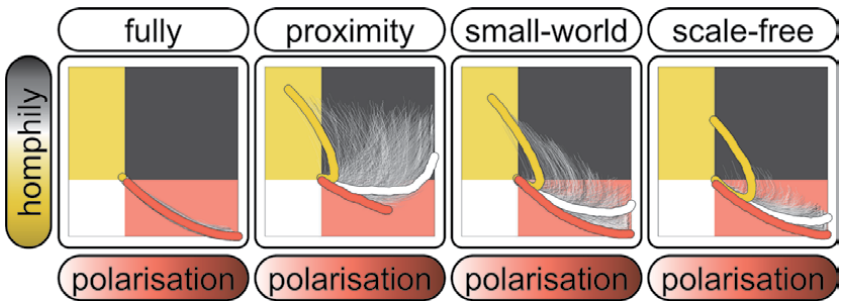


Figure 4.
Trajectories of social polarisation (horizontal axis) and homophily (vertical axis) simulated in four different network topologies. Each realisation for some persuasiveness, θ and opinion volatility μ is marked with a curve. All curves or realisations have a nearby starting point, which marks the polarisation and homophily of a random distribution of opinions. For each topology, the four quadrants with a higher (or lower) polarisation and a higher (or lower) homophily are coloured and the three trajectories with the highest and lowest polarisation and the highest homophily are marked with thick curves.

For some of the trajectories, it is observed that the first few rounds of propaganda increase the polarisation and decrease the homophily. After many rounds of propaganda, the level of homophily might increase, indicating the formation of clusters of nodes with similar opinions, particularly on a proximity network. In some cases, polarisation might be decreased, but only after homophily has decreased (and not the other way around), meaning that first, the observed changes in opinion dynamics happen at a local level and then, they might be perceived at a global scale. Notice, however, that very few trajectories reach less polarisation than their starting point. Thus, propaganda or similar external forces tend to increase polarisation and frequently will produce a higher level of polarisation than the one observed with a random distribution of opinions.

3.1 Parameter space

The observed levels of polarisation and homophily depend on the persuasiveness of the propaganda θ, and the opinion volatility μ. On a proximity network, for instance, with highly persuasive propaganda ($\theta \approx 1$) and volatile opinions ($\mu \approx 1$) after only a few rounds of propaganda, there is a highly polarised society, with highly homophilic interactions. However, if propaganda is not as seductive or if individuals do not update their views easily, it takes several rounds of propaganda to observe a polarised society (**Figure 5**).

For some values of θ and μ, there is extreme sensitivity to the parameters. On a proximity network, with higher values of the persuasiveness of propaganda θ,

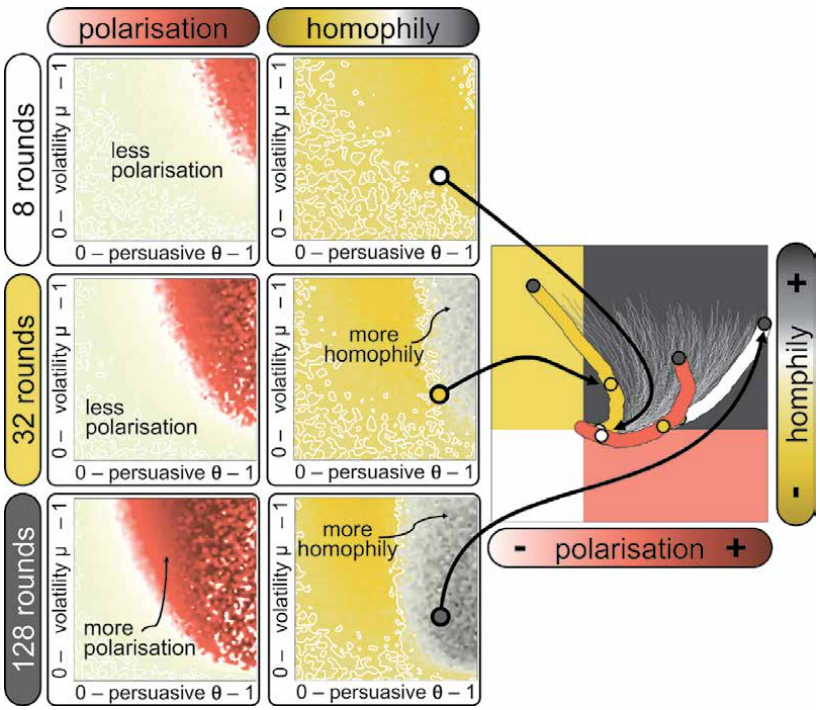


Figure 5.
Observed levels of polarisation (left) and homophily (middle) on a proximity network according to some values of the persuasiveness of propaganda θ (horizontal axis) and the volatility of opinions μ (vertical axis) after 8, 32 and 128 rounds of propaganda. Higher levels of polarisation and homophily are darker, representing extreme views and a homophilic society respectively, and lower levels of polarisation and homophily are lighter, representing consensus and frequent exchanges between people with different views. For the same values of $\theta = 0.85$ and $\mu = 0.35$, with the same initial (random) opinions, 250 realisations of the dynamics follow different trajectories (right), where the levels of polarisation and homophily after 8, 32 and 128 rounds of propaganda are highlighted.

society might be alternatively highly polarised or close to a consensus after many rounds of propaganda with very small changes in the two parameters. Furthermore, even with the same initial opinions and with the same values of θ and μ, society might reach very different levels of polarisation and homophily (right part of **Figure 5**). Individuals exposed to propaganda are randomly picked and according to their opinion, they might be seduced by it and share it with their contacts, or ignore it, thus, altering the outcome after that round of propaganda. With only a few waves of propaganda, the outcome might be similar, but those small changes are cumulative and so after many rounds, the outcome might be a society close to extremism or even close to consensus, even if the starting point is the same.

The first rounds of propaganda decrease the homophily of society so that people with some extreme view have frequent interactions with others with different views. As the number of propaganda rounds evolves, opinion clusters are formed, and so interactions become more and more frequent between individuals with similar views. Thus, even if at a global scale the level of polarisation is increasing, after many rounds of propaganda, people might be less aware of the existence and abundance of different views. Extreme opinions might become more frequent because of propaganda. A similar -although less pronounced- polarising and homophilic society might be frequently observed on a scale-free and a small-world network, although the presence of hubs and shortcuts in the network reduces the creation of opinion clusters (**Figure 6**).

The fully-connected network helps to observe the dynamics of opinions without any relevant network structure. With some level of persuasiveness θ, and opinion volatility μ, society eventually reach polarisation. With more rounds of propaganda, polarisation increases up to extremism, and only with no persuasiveness ($\theta = 0$) or no volatility ($\mu = 0$) society remains without extremism. However, for different network topologies, propaganda might have a different impact. Particularly in the case of a proximity network (with high values of θ) and in the case of a scale-free network (with medium values of θ) propaganda might increase homophily and in some cases, reduce polarisation.

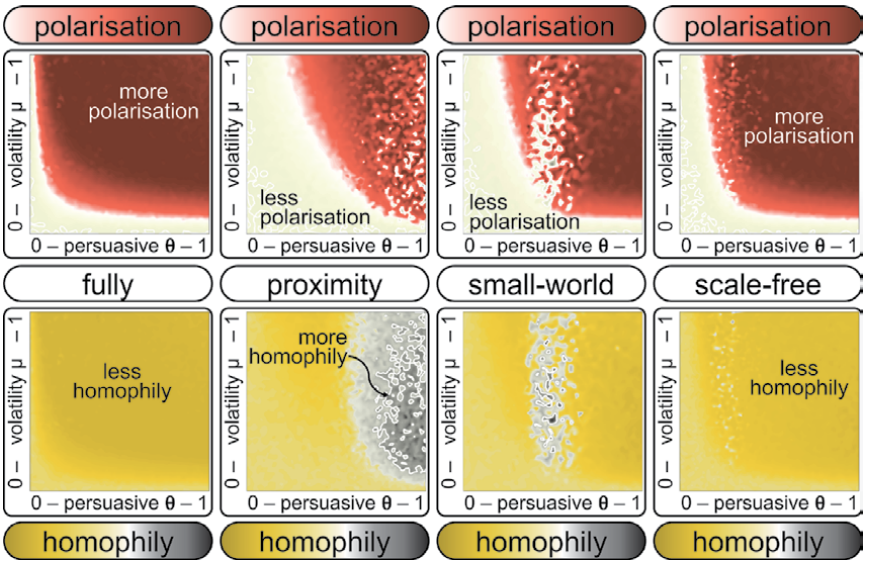


Figure 6.
Observed levels of polarisation (top) and homophily (bottom) according to some values of the persuasiveness of propaganda θ (horizontal axis) and the volatility of opinions μ (vertical axis) after 128 rounds of propaganda. Four network topologies are considered, a fully-connected, a proximity, a small-world and a scale-free network from left to right.

4. Conclusions

Social models are a simplification of very complex processes which happen at an individual level but might be able to capture some collective emergent aspects. In terms of opinion dynamics, modelling individual views as a number, simplifying external forces such as propaganda, simulating interactions and a process of opinion updating let us detect emergent patterns, including an increase in the global levels of polarisation and the frequency of homophilic interactions between individuals.

The network structure plays a significant role, as the emergence of homophilic clusters which reinforce their opinions is detected, particularly on a network where there is a large distance between nodes, such as a proximity network.

The observed results in terms of the trajectories and the observed levels of polarisation and homophily after many rounds of propaganda show that there might be a high sensitivity concerning the parameters. Two simulations under the same network structure and even the same initial opinions and parameters might follow different trajectories and end with substantially distinct levels of homophily and polarisation. The model initially exposes 1% of the population to some propaganda and depending on who is exposed, the dynamic changes and eventually reach very different states. For some regions in the parameter space, there is unpredictability in the state in which society will be after propaganda.

In the simulated networks, the average degree is 7.6 for the proximity network and 10 for the small-world and the scale-free network. The intensity of interactions, measured as the degree of the nodes, accelerates or frictions the diffusion of propaganda, and thus, accelerates of frictions polarisation and homophily as well. A less-connected society is more prone to the creation of homophilic clusters.

4.1 What is different between a highly polarised society and one with little polarisation

On a highly polarised society, individuals become “immune” to propaganda which does not support their views and dismiss it easily, whereas propaganda which supports their views is confirmation of their beliefs and takes individuals into even more extreme and plarised views. On a polarised society, even with little levels of homophily (meaning that individuals are likely to be exposed to both types of propaganda), individuals are eventually too biased in favour of any of the extreme views, which becomes too difficult to change.

On a society with little levels of polarisation, views could either have a consensus on one of the two extremes, in which case, propaganda in favour of any of the opinions has little impact. This case happens when one of the two views becomes dominant at early stages, in which case, individuals also become “immune” to propaganda (and since the first propaganda they are exposed is $+1$, that view is slightly more likely to become dominant in the long run).

However, the most frequently observed consensus is one in which barely anyone has extreme views, propaganda in favour of the two views flows between most individuals and they update their opinion accordingly, but not enough to reject further waves of propaganda and keep updating their opinion.

Acknowledgements

This chapter was completed with support from the PEAK Urban programme, funded by UKRI’s Global Challenge Research Fund, Grant Ref: ES/P011055/1.

Conflict of interest

The author declare no conflict of interest.

Author details

Rafael Prieto Curiel
Centre for Advanced Spatial Analysis, University College London, London, UK

*Address all correspondence to: rafael.prieto.13@ucl.ac.uk

References

[1] Jeffrey H Johnson. The future of the social sciences and humanities in the science of complex systems. *Innovation–The European Journal of Social Science Research*, 23(2):115–134, 2010.

[2] Julian CR Hunt, Yulia Timoshkina, Peter J Baudains, and Steven R Bishop. System dynamics applied to operations and policy decisions. *European Review*, 20(3):324–342, 2012.

[3] Jeffrey H Johnson. The "can you trust it?" problem of simulation science in the design of socio-technical systems. *Complexity*, 6(2):34–40, 2000.

[4] Eric Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99 (suppl 3):7280–7287, 2002.

[5] Xiaoshan Pan, Charles S Han, Ken Dauber, and Kincho H Law. A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *Ai & Society*, 22(2):113–132, 2007.

[6] Shane D Johnson and Elizabeth R Groff. Strengthening theoretical testing in criminology using agent-based modeling. *Journal of Research in Crime and Delinquency*, 51(4):509–525, 2014.

[7] Serge Galam, Yuval Gefen, and Yonathan Shapir. Sociophysics: A new approach of sociological collective behaviour. i. mean-behaviour description of a strike. *Journal of Mathematical Sociology*, 9(1):1–13, 1982.

[8] Martin Short, Jeffrey Brantingham, Andrea Bertozzi, and George Tita. Dissipation and displacement of hotspots in reaction-diffusion models of crime. *Proceedings of the National Academy of Sciences*, 107(1):3961–3965, 2010.

[9] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684): 440–442, 1998.

[10] Dirk Helbing, Dirk Brockmann, Thomas Chadefaux, Karsten Donnay, Ulf Blanke, Olivia Woolley-Meza, Mehdi Moussaid, Anders Johansson, Jens Krause, Sebastian Schutte, and Matjaž Perc. Saving human lives: What complexity science and information systems can contribute. *Journal of Statistical Physics*, 158(3):735–781, 2015.

[11] Gérard Weisbuch, Guillaume Deffuant, Frédéric Amblard, and Jean-Pierre Nadal. Meet, discuss, and segregate! *Complexity*, 7(3):55–63, 2002.

[12] Rafael Prieto Curiel and Humberto González Ramírez (2021). Vaccination strategies against COVID-19 and the diffusion of anti-vaccination views. Scientific Reports, 11(1).

[13] Bill Bishop. *The big sort: Why the clustering of like-minded America is tearing us apart*. Houghton Mifflin Harcourt, 2009.

[14] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3): 554–559, 2016.

[15] Vítor V. Vasconcelos, Simon A Levin, and Flávio L Pinheiro. Consensus and polarization in competing complex contagion processes. *Journal of the Royal Society Interface*, 16(155):20190196, 2019.

[16] Kazutoshi Sasahara, Wen Chen, Hao Peng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. On the inevitability of online

echo chambers. *arXiv preprint arXiv: 1905.03919*, 2019.

[17] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[18] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[19] Seth Flaxman, Sharad Goel, and Justin M Rao. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1):298–320, 2016.

[20] Gilat Levy and Ronny Razin. Social media and political polarisation. *LSE Public Policy Review*, 1(1), 2020.

[21] Pennycook, G., Cannon, T. D., and Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. Journal of Experimental Psychology: General, 147(12), 1865.

[22] Dustin P Calvillo, Bryan J Ross, Ryan JB Garcia, Thomas J Smelter, and Abraham M Rutchick. Political ideology predicts perceptions of the threat of COVID-19 (and susceptibility to fake news about it). *Social Psychological and Personality Science*, page 1948550620940539, 2020.

[23] Noah E Friedkin and Eugene C Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3–4):193–206, 1990.

[24] J Scott Brennen, Felix Simon, Philip N Howard, and Rasmus Kleis Nielsen. Types, sources, and claims of COVID-19 misinformation. *Reuters Institute*, 7:3–1, 2020.

[25] Larson, H. J. (2020). A lack of information can become misinformation. Nature, 306-306.

[26] Oberiri Destiny Apuke and Bahiyah Omar. Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, page 101475, 2020.

[27] Bertram Düring, Peter Markowich, Jan-Frederik Pietschmann, and Marie-Therese Wolfram. Boltzmann and Fokker–Planck equations modelling opinion formation in the presence of strong leaders. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 465 (2112):3687–3708, 2009.

[28] Bertram Düring and Marie-Therese Wolfram. Opinion dynamics: inhomogeneous Boltzmann-type equations modelling opinion leadership and political segregation. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 471(2182), 2015.

[29] Rafael Prieto Curiel and Steven Richard Bishop. Modelling the fear of crime. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 473(2203), 2017.

[30] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591, 2009.

[31] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(4):87–98, 2000.

[32] Luís M.A. Bettencourt, Ariel Cintrón-Arias, David I. Kaiser, and Carlos Castillo-Chávez. The power of a good idea: quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications*, 364:513–536, 2006.

[33] Deepak Bhat and S Redner. Polarization and consensus by opposing

external sources. *Journal of Statistical Mechanics: Theory and Experiment*, 2020 (1):013402, 2020.

[34] Janusz Holyst, Krzysztof Kacperski, and Frank Schweitzer. Social impact models of opinion dynamics. *Annual Reviews of Computational Physics*, 9:253–273, 2002.

[35] Krzysztof Kacperski and Janusz Holyst. Opinion formation model with strong leader and external impact: a mean field approach. *Physica A: Statistical Mechanics and its Applications*, 269(2):511–526, 1999.

Chapter 3

Exploring Links between Complexity Constructs and Children's Knowledge Formation: Implications for Science Learning

Michael J. Drobonikuˎ Heidi Kloosˎ Dieter Vanderelst and Blair Eberhart

Abstract

This essay brings together two lines of work—that of children's cognition and that of complexity science. These two lines of work have been linked repeatedly in the past, including in the field of science education. Nevertheless, questions remain about how complexity constructs can be used to support children's learning. This uncertainty is particularly troublesome given the ongoing controversy about how to promote children's understanding of scientifically valid insights. We therefore seek to specify the knowledge–complexity link systematically. Our approach started with a preliminary step—namely, to consider issues of knowledge formation separately from issues of complexity. To this end, we defined central characteristics of knowledge formation (without considerations of complexity), and we defined central characteristics of complex systems (without considerations of cognition). This preliminary step allowed us to systematically explore the degree of alignment between these two lists of characteristics. The outcome of this analysis revealed a close correspondence between knowledge truisms and complexity constructs, though to various degrees. Equipped with this insight, we derive complexity answers to open questions relevant to science learning.

Keywords: cognitive development, science education, conceptual change, complex adaptive systems, thermodynamics, interdisciplinary theory

1. Introduction

> *We need to move toward a systems view that describes scientific concepts as complex.*
>
> *– Andrea A. diSessa [1]*

It has long been accepted that children's knowledge formation defies straightforward processes of passive attention and associative learning [2, 3]. For example, rather than absorbing information indiscriminately, children will actively seek out some aspects of information, while ignoring others. Children can also imagine alternative realities, even fantasies that lack a grounding in reality [4]. This poses a problem when it comes to the question of how to improve children's knowledge. For example, it

remains unclear how to support the learning of abstract science concepts, especially when children hold incorrect naïve beliefs about the pertinent science phenomenon [5]. In the current paper, we seek to contribute to this conversation by systematically exploring links between knowledge formation and complexity constructs.

In order to offer a relatively unbiased discussion of the complexity of knowledge, we first identified central truisms about knowledge formation that are broadly supported by the literature. We then provided a glossary of complexity constructs that are potentially useful in understanding knowledge formation. Equipped with these two lists, we then evaluated whether facts about knowledge formation are anticipated by complexity constructs. In turn, this cross-tabulation served as a theoretical anchor to derive answers from complexity to open questions on children's science learning.

2. Established insights about knowledge formation

Picture a child trying to balance a beam on a fulcrum. The principle of physics that matters in this task is that of weight distribution in the beam. While children are capable of detecting the beam's weight distribution, they sometimes focus on the beam's visual symmetry instead. The result is that children have trouble balancing beams with asymmetrical weight distribution; they try to balance them at their geometric center instead of their center of mass. This finding illustrates established facts about (1) the nature of knowledge, (2) the process of knowledge acquisition, and (3) the process by which knowledge is changed (see **Table 1** for an overview).

2.1 Nature of knowledge

A child who insists that a beam should balance at the geometric center is said to hold the mistaken belief that objects balance in the middle [6]. The nature of such a belief (or knowledge, more generally) is necessarily elusive, as it cannot be seen directly. For this reason, numerous models of knowledge have been proposed to rectify phenomenological and empirical findings (e.g., [7, 8]). Considered in the aggregate, the models largely agree on two characteristics of knowledge: (i) that knowledge is organized into structures, and (ii) that there are different kinds of knowledge structures. We elaborate on each of these characteristics next.

2.1.1 Truism 1: Knowledge is organized into structures

Rather than existing as encapsulated factoids, knowledge consists of interlined representations of experiences, also referred to as *schemas* or *mental models* [9–11]. Early evidence for such knowledge organization came from children's systematic errors in Piaget's classical conservation tasks [12]. Children spontaneously and consistently honed in on a particular variable to respond, suggesting the presence of mental structures that make one variable more salient than another. Numerous additional examples stem from errors in categorization tasks [13], causal-reasoning tasks [14], and learning tasks [15]. They suggest that knowledge needs to be conceptualized as an ordered set. A child's belief about beam-balancing is an example of such a knowledge structure.

2.1.2 Truism 2: There are different types of knowledge structures

Agreement also exists that there are qualitative differences in knowledge organization. A prominent distinction is between *implicit* and *explicit* knowledge: Only

Nature of Knowledge	
Structure	Knowledge is organized coherently, rather than consisting of incoherent bits.
Diversity	Knowledge differs in various aspects, including implicit versus explicit, superficial versus deep, or narrow versus broad.
Acquisition of Knowledge	
Holistic construal	Knowledge is actively construed (abducted), rather than being a direct reflection of experiences.
Context dependence	The details of knowledge depend on various contextual factors, including the social context, the nature of specific tasks, and the available tools.
Change of Knowledge	
Persistence	Mistaken beliefs often persist, even to the point of affecting perception, despite their obvious shortcomings.
Role of conflict	Presenting children with the shortcomings of their naïve beliefs creates a conflict that can lead to conceptual change.

Table 1.
Central characteristics of knowledge formation.

explicit, not implicit knowledge, can be reported on [16, 17]. Another example is the distinction between *surface knowledge* and *deep knowledge*: Only deep knowledge, not surface knowledge, can be transferred to new situations [18, 19]. And yet another example is the distinction between *preconceptions* and *misconceptions* (e.g., [20]): While both types of knowledge lead to mistaken performance, only misconceptions, not preconceptions, persist [21]. These and other distinctions proved useful in capturing unexpected behavior, including that of balance-beam tasks [22].

2.2 Acquisition of knowledge

The child presented with a balance-beam task will eventually realize the relevance of weight distribution and succeed in balancing off-center beams. That is to say, the child will eventually learn. This process of learning, like knowledge itself, cannot be seen directly [23]. Sure enough, there are numerous open questions and disagreements about how to best describe the process by which knowledge is formed [24]. There are, however, two characteristics of learning that are broadly agreed upon: (i) that knowledge is construed through the child's activity, and (ii) that aspects of the context strongly affect what is being learned. We elaborate on each of these characteristics next.

2.2.1 Truism 3: Knowledge is construed

At first glance, knowledge appears to reflect outside information, as if outside information was transported into the mind directly. There is indeed suggestive evidence in support of such passive learning [25]. On the other hand, however, there is widespread agreement that learning requires an active mind. Piaget coined the term 'constructivism' to capture this idea: The mind, rather than passively soaking up information, must actively build knowledge. As a result of such construal, knowledge structures might come into existence nonlinearly, reflected in the aha-moment of sense-making (see also *abduction*; [26]). DiSessa [1] captured this nonlinearity in the proposed trajectory from a naïve learner to the conceptually competent individual (see **Figure 1** for a schematic illustration of the suggested nonlinearity).

2.2.2 Truism 4: Knowledge formation depends on the context

Evidence suggests that learning is strongly dependent on the context, even when the context appears irrelevant to the specifics of what needs to be learned. Illustrative evidence of such context dependence can again be found with Piaget's conservation tasks: Despite showing robust performance in the classical version of the task, the mere number of times children were asked for a comparative judgment affected their performance (e.g., [27]). Evidence also comes from the balance-beam task: Children managed to balance the beams better with their eyes closed than with their eyes open [6]. Overall, context (e.g., cultural, societal, physical) has the ability to influence how the information ends up being utilized within the system of knowledge [28, 29].

2.3 Change of knowledge

A child presented with the balance-beam task is unlikely to enter the situation without prior knowledge. It might pertain to general ideas about what to expect, or it can pertain to very specific ideas about how to solve a task (e.g., the belief that beams balance at their center). For science learning to take place, incorrect prior knowledge has to be replaced, a process also known as *conceptual change* [30]. Exactly how to promote conceptual change remains a challenge in science education [31]. At the same time, there are two characteristics of conceptual change that are broadly acknowledged: (i) that existing knowledge structures have a strong tendency to persist, and (ii) that the experience of conflict can prompt conceptual change. We elaborate on each of these characteristics next.

2.3.1 Truism 5: Knowledge structures resist change

There is wide-spread agreement that existing knowledge structures can persist despite strategic changes in the learning context. The domain of science learning is packed with examples of such persistence of mistaken beliefs [32]. It appears as though existing knowledge can affect how one perceives the surroundings, even to the point of inventing improbable experiences (e.g., [33]). The example of beam-balancing illustrates this peculiarity: It is as if children's beliefs about beam-balancing impedes their ability to take in conflicting experiences. In fact, Karmiloff-Smith and Inhelder [6] described children who actively ignored the evidence of a beam tipping over when they attempted to balance it in the middle.

2.3.2 Truism 6: Perceived conflict facilitates conceptual change

Conceptual change is possible when a pedagogy is used that highlights the shortcomings of the existing belief [34, 35]. The power of conflict can be traced to the work of Piaget [36], Festinger [37], and Dewey [38]. The argument is that perceived contractions generate conceptual conflict, which, in turn, serves as a catalyst for deeper forms of cognitive processing [39]. Presumably, children who hold a naïve belief about beam-balancing can experience conflict as they continue to play with the beams, which, in turn, might prompt them to replace their naïve belief.

2.4 Summary of central characteristics of knowledge formation

In the first part of this preliminary section, we sought to systematize the vast literature on knowledge formation in a way that highlights central characteristics of this process. On the question of the nature of knowledge, we honed in on the

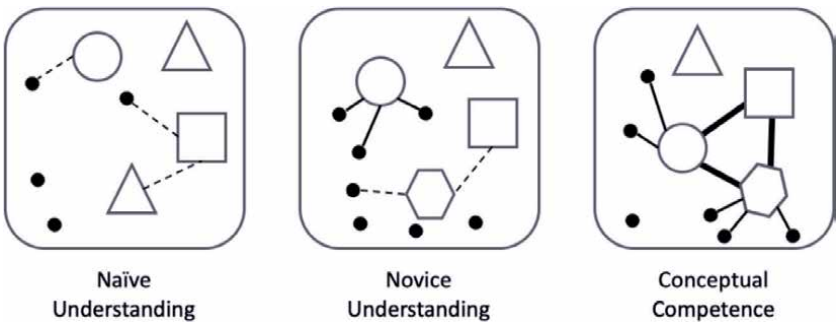


Figure 1.
Proposed illustration of knowledge at three stages of the learning process (adapted from [1]). Shapes are thought to be exemplars of experiences represented in the mind. They become organized as conceptually competent knowledge develops. Complexity constructs provide further suggestions to consider.

ideas that knowledge is organized (rather than existing as an isolated fixture) and that several distinct types of organization exist (rather than differing merely in content). On the question of learning, we honed in on the ideas that knowledge emerges via the active involvement of the learner (as opposed to being transmitted passively) and that learning is affected by the context, whether relevant or not (rather than be affected merely by what matters most). On the question of conceptual change, we highlighted the persistence of mistaken beliefs and the power of conflict to prompt conceptual change.

3. Glossary of central complexity constructs

There are excellent sources available to introduce complexity science (e.g., [40–44]). The field of complexity science can nevertheless appear unorganized, featuring constructs that are not fully integrated with each other. It is not immediately obvious, for example, how constructs such as attractors, scale-free patterns, or synchrony relate to one another (or differ from each other, for that matter). This hinders progress on how complexity theory could help with knowledge formation. For this reason, we provide a review of selected complexity constructs. We have organized the list by the type of system that best exemplifies the selected constructs: (1) non-living systems, (2) living systems, and (3) thermodynamic systems (see **Table 2** for an overview).

3.1 Constructs from the study of non-living systems

There are several non-living systems that have been used as model domains to explore complex systems, including cellular automata [45], oscillators [46], or electricity grids [47]. Common to all of these systems is that their elements interact with each other. The nature of this interaction is fixed, as is the nature of the elements in these systems. Yet, despite this simplicity, non-living systems can behave in complex ways. Constructs that have been explored in these systems include self-organization, chaos, hysteresis, attractors, autocatalysis, self-organized criticality, and scale-free patterns. We describe these constructs next.

3.1.1 Self-organization

Arguably at the heart of complexity science, self-organization is the process by which global patterns form through local interaction of the system's elements

Constructs from Non-Living Systems	Brief Definition	Living Systems	Thermo-dynamics
Self-organization	The emergence of spatiotemporal patterns through the interaction of system elements.	**X**	**X**
Chaos	Behavior is highly sensitive to initial conditions because of the amplification of interacting constraints.	**X**	**X**
Hysteresis	A nonlinear shift takes place at a moment in time that is affected by the cumulative history of the system.	**X**	
Attractors	A behavior toward which the system navigates.	**X**	
Self-organized criticality	A state of the system in which several behavioral options are available.	**X**	
Self-similarity (e.g., scale-free patterns, pink noise, fractals)	Patterns are composed of elements that look similar or identical to the patterns they make up.	**X**	
Constructs from Living Systems	**Brief Definition**	**Non-Living Systems**	**Thermo-dynamics**
Affordance	Sense-making of the surrounding depends on the action of the individual.		
Synchrony	System elements mutually constrain each other as they interact in a circular way.	**X**	
Self-preservation (e.g., autopoiesis, centripetality)	The system carries out processes that contribute to its own self-maintenance.		**X**
Constructs from Thermodynamic Systems	**Brief Definition**	**Non-Living Systems**	**Living Systems**
Balance/Equilibrium	The system settles on an organization that is most probable given the existing distribution of energy.		
Dissipation pressure	System elements organize themselves into patterns to dissipate the gradient established by energy clusters.		
Autocatakinetics	System elements become increasingly more organized in the service of the dissipation pressure.		**X**
Teleodynamics	The coming together of mutually constraining processes that perpetuate each other, seemingly bestowing agency to structures.	**X**	**X**

Note: While the complexity constructs are listed under only one type of system, they apply to other systems as well (marked by X in the last two columns of the table).

Table 2.
Overview of selected complexity constructs, separated by type of system that exemplifies them best.

[48]. When ordered structures are caused by self-organization, there is no blueprint or central control. Instead, the observed pattern is an emergent property [49]. The marking of sand dunes is an example of such self-organization. It stems from the "interplay of windborne transport, collision-driven piling up, and slope-shaving avalanches" ([50], p. 1084). Another example is the synchronization of adjacent metronomes that are initially out of sync. Eventually, the metronomes settle on a synchronized rhythm by virtue of sharing the surface they are placed on [51]. In each of these cases, the interaction among individual elements gives rise to overarching patterns that could be reduced neither to the elements nor the outside.

3.1.2 Chaos

In chaotic systems, future behavior is sensitive to the initial conditions [52]. Chaos can be illustrated with the butterfly effect as a metaphor: A butterfly fluttering its wings over a flower in China can, in principle, cause a hurricane in the Caribbean [53]. A simple system that exhibits chaotic behavior is the double pendulum: Small differences in the initial angles of the pendulum arms are amplified several orders of magnitude in the course of just a few seconds [54]. Chaotic behavior is the result of the coming together of various factors that allow a change to become amplified (or dampened) as the change reverberates through the system. The result is unpredictable behavior of the system, despite having fully deterministic links among its individual elements.

3.1.3 Hysteresis

Hysteresis describes a sudden change in behavior that is modulated by the system's history. Relevant here is the direction in which an outside parameter changes (from low to high, or from high to low). A thermostat provides an illustrative example of this phenomenon: Its function is to detect the temperature of the surrounding to control whether the heat should be on or off. Importantly, the change in the system's on–off status is not necessarily the result of an absolute outside temperature. Instead, the thermostat might have a different temperature threshold for switching the heating on than for switching it off [55]. This allows the thermostat to avoid repeatedly switching the heating on and off when the temperature hovers around the set point. The mathematical branch of *catastrophe theory* provides further specifications of the patterns of hysteresis, including how the presence of an additional outside parameter can modulate hysteresis (see also *cusp-catastrophe*; [56]).

3.1.4 Attractors

An attractor is a state to which the system returns after having been perturbed away from it. Attractors come in several forms, the simplest of which is a point attractor. Consider, for example, a damped harmonic oscillator. The behavior of the oscillator depends on its mass, the spring stiffness, and the damping coefficient—all of which are referred to as control parameters. These parameters determine the details of the oscillator's resting states. If the oscillator were to be pushed away from its resting state, it will eventually return to it, thus demonstrating the state as an attractor for the system [57]. Other forms of attractors are periodic attractors (i.e., the cyclical moving through several stable states; *limit cycle*) and strange attractors (i.e., the non-periodic or chaotic movement through several states).

3.1.5 Self-organized criticality

Self-organized criticality combines the ideas of self-organization and attractors, stating that systems maneuver themselves into a specific state, referred to as the *critical state* [58]. In a critical state, small perturbations can lead to large-scale or catastrophic changes in the system (e.g., [59]). Bak and Chen [59] proposed that the systems attracted to a critical state exhibit $1/f$ noise. The spectral density of the system's response to perturbation can be approximated as: $D_f \approx 1/f^{\alpha}$ (with $0.50 < \alpha < 1.50$). A well-studied example includes earthquakes, both simulated and real ones.

3.1.6 Self-similarity

Self-similarity is present when the elements the elements resemble the very pattern that they make up [60]. The geometric shape known as the Sierpiński triangle is a famous example: Upon zooming in, the parts of the triangle resemble the triangle itself. The relevance of self-similarity lies in the relation among hierarchically nested patterns. In a self-similar pattern, there are no unique 'starter' elements, as each element is itself composed of entire patterns. That is to say, there is no characteristic scale at which the behavior of a system resides, an idea captured in *scale-free patterns* (see also *cumulative advantage*; [61, 62]). Self-similar patterns are relevant in the understanding of *fractals* and *power-law* distributions, also referred to as *pink noise*. Common to these terms is the idea that there is a long-range dependence among the different levels of organization in a system.

3.2 Constructs from the study of living systems

Like non-living systems, living systems consist of interacting elements that give rise to patterns of organization. Obvious examples include systems of individual animals (e.g., a school of fish, a flock of birds, an ant hill, a group of synchronizing fireflies) or of entire species (e.g., ecosystem). There are also systems within an individual animal, like when cells organize into an organ system [63, 64]. Given the interaction among elements, all of the complexity constructs identified for non-living systems apply here as well. For example, the organizations observed in these systems (e.g., nest building, foraging routes, behavior of crowds) stem from processes of self-organization. There is also evidence of hysteresis (e.g., the switch from fight to flight) and the presence of self-similar patterns (e.g., the branching of trees).

There is, however, a crucial difference between living and non-living systems: Rather than being fixed, elements in a living system can change (see also *complex adaptive systems* vs. *complex physical systems*; [41]). In other words, "living" elements can learn, adapt, and evolve, which, in turn, changes the relation they have to each other. In an ecological niche, for example, entirely new elements can appear (e.g., a new individual in a group), yielding new interactions and configurations. For this reason, some complexity constructs pertain only to living systems. We consider the constructs of affordance, synchrony, and self-preservation.

3.2.1 Affordance

An affordance is the opportunity for action that is made possible by the environment. The construct was developed by James Gibson as an explanation to how animals make sense of and navigate their surroundings [65]. An example of an affordance is the optic flow, a vector field of the perceived motion of static objects

that is established through the movement of an animal. The optic flow does not exist entirely in the surrounding, nor is it a process of internal mental symbol manipulation. Instead, it is caused by the relative motion between an agent and the scene. Many insects have visual systems that are specialized for extracting optic flow. For example, a bee flying through a tapering corridor would experience an increase in translational flow as the corridor narrows, unless the bee slows down [66].

3.2.2 Synchrony

Synchrony refers to the coordination that takes place among the elements of a system (see also *circularity*, *interdependence*, *coupling*). While it can be found in non-living systems (e.g., coupled metronomes), it has been studied extensively in living systems, including in the behavior of molecules, plants, animals, neurons, muscles, bodily regulations, and human relations [67, 68]. There is, in fact, an entire subfield of mathematics focused on theories related to synchrony—namely, to capture the degree to which elements affect each other's behavior in interdependent ways (see also *coupling strength*). When a system is tightly coupled, its elements coordinate closely with each other. In contrast, when a system is loosely coupled, its elements have little to no effect on each other.

3.2.3 Self-preservation

Living systems appear to perpetuate their own organization autonomously, what Darwin famously referred to as a "struggle for existence" [69]. There are a number of complexity constructs that can be used to describe this process of self-preservation. The concept of *agency*, for example, captures the tendency to act on one's own behalf, thus contributing to a system's ability to maintain itself [70]. The concept of *autopoiesis* is another example of self-preservation. An example is the process by which the cells of an organism are able to reproduce and maintain themselves via the production of and interaction between individual elements [71]. Some autopoietic systems can even undergo *recursive self-maintenance* in which the agent is able to select from a variety of processes, depending on their environmental circumstances [72]. Yet another construct that captures self-preservation is that of *centripetality*. This refers to a system's capacity to produce and maintain its own complexity by attracting resources into its circular patterns of self-organization [73].

3.3 Constructs from the study of thermodynamic systems

A third set of complexity constructs stems from thermodynamic systems—systems that illustrate the laws of thermodynamics [74–76]. These systems consist of an energy source, a set of elements that are sensitive to the outside energy source, and a mutually constraining coupling among elements. An illustrative example is a pot of water placed on a burner: The heat from the burner constitutes the energy source; the water molecules are the elements (sensitive to the heat); and the push–pull movement among the water molecules captures their coupling strengths. Another example is an ecosystem [77]: The resources available in the surrounding constitute the energy source; the species of the ecosystem are the elements (sensitive to these resources); and the relations among the species (predator–prey; symbiotic) capture their coupling strength. Relevant constructs from these systems are that of balance, gradient dissipation, autocatakinetics, and teleodynamics. We describe these next.

3.3.1 Balance

Thermodynamic systems move toward a state in which forces are balanced (also referred to as *homeostasis* or *equilibrium*). Grounded in fundamental laws of physics, balance exists when there is no longer any net change in forces, influences, and/or reactions. In that sense, thermodynamics offers a traceable endpoint to behavior (a purpose, so to speak), namely, in achieving balance. Outside of physics, balance is also used to indicate steady or stationary conditions in branches such as evolution, economy, and social sciences [78]. An example of balance is captured in the term of *ascendancy*, which is the degree of relative stability in an ecosystem, shown to increase over evolutionary timescales [79–81].

3.3.2 Dissipation pressure

In addition to endowing systems with the purpose of reaching a balance, thermodynamics also identifies the conditions necessary for systems to do so: The push toward balance comes from the presence of clustered energy. This is because the presence of clustered energy, in addition to affecting the system, also sets up a gradient that needs to be dissipated (captured in the second law of thermodynamics; [82]). For example, the mere presence of clustered heat in a cup of tea sets up a gradient to be dissipated (i.e., the heat clustered in the cup will eventually disperse to reach thermal equilibrium). This pressure to dissipate an energy gradient can push the system to create micro-clusters of energy. In boiling water, for example, water molecules organize themselves into vapor pockets that contain some of the heat (see also *morphodynamics*; [83]). Put differently, the pressure to dissipate an energy gradient provides opportunities for the system to organize itself (see also *antifragility*; [84]).

3.3.3 Autocatakinetics

Under some circumstances, systems become increasingly more ordered, seemingly going against the push for dissipation of clustered energy. Animals and plants, for example, appear to pursue the survival of their species, coming up with increasingly more efficient ways to harness and retain resources. These systems are known to be autocatakinetic [85]. **Figure 2**, adapted from Swenson [85], illustrates how the emergence of progressively more organized forms of a system is possible under the law of maximum entropy production. An external energy source (i.e., one that is outside of a local, open system) clusters to create an energy gradient that must be dissipated in order to reach entropic balance in the broader (closed) global system. In moving toward dissipation, a second cluster of energy emerges in the local system, composed of the self-organized behavior of the system's elements. This energy cluster, in turn, defines another energy gradient, hence another push toward dissipation that contributes to the entropic balance of the global system.

3.3.4 Teleodynamics

Teleodynamics is yet another principle that seeks to explain how elements of a system become increasingly more ordered, despite the push toward maximum entropy [83, 86]. The idea is that order is perpetuated when mutually supporting processes come together. A so-called *autocell* (or *autogen*) is a model system that can illustrate this idea. This model is based on two processes, that of autocatalysis (i.e., the mutual facilitation of two or more chemical reactions) and that of containment (i.e., the forming of enclosures from the biproduct of the autocatalytic reactions). The interaction of these two processes (i.e., autocatalysis and containment) allows

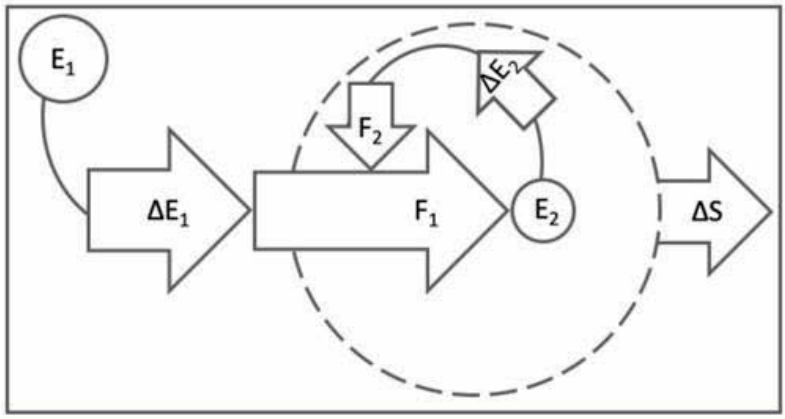


Figure 2.
Illustration of autocatakinetic closure, adapted from Swenson [85]. The solid frame defines the boundary of a global (closed) system. The dashed circle defines the boundary of a local (open) system within the global one. The energy source E1 defines an energy gradient (ΔE1) that needs to be dissipated (F1) to reach entropic balance (ΔS). In moving toward dissipation, a second cluster of energy emerges (E2), which consists of the self-organized behavior of the system's elements. This second energy cluster, in turn, defines an energy gradient (ΔE2) and, thus, another push toward dissipation (F2).

each of them to continue, even as reactants are used up and the enclosures break apart. The outcome is a self-repair and self-replication of sorts (also see *hypercycles*, *autogenesis*, *negentropy ratchet*; [87–89]).

3.4 Summary of central complexity constructs

In the second part of this preliminary section, we sought to review central complexity constructs in a way that facilitates the attempted link between complexity and knowledge formation. In total, we selected over a dozen complexity constructs, some of which apply to all systems (e.g., self-organization, attractors), and some of which apply to some systems exclusively (e.g., agency, hysteresis). For each of these terms, we offered an explanation at the level of phenomenology, bypassing mathematical advances. Emphasis was placed on providing a general sense of the concepts with explanations that were broad enough to subsume several complexity constructs (e.g., synchrony vs. coordination).

4. Cross-tabulation of knowledge and complexity

The link between cognition and complexity is invoked often, as the quote at the top of the paper suggests (see also [90–94]). However, it is not always clear if the ideas are applied consistently, as neither the field of cognition nor the field of complexity is straightforward. Having provided an organization of both areas (Sections 2 and 3 above), we are in the position to address the link systematically. **Table 3** provides an overview of our cross-tabulation.

4.1 Complexity links to the truisms of knowledge formation

4.1.1 Link 1: Complexity in the structural organization of knowledge

There are several complexity constructs that anticipate knowledge being organized. *Self-organization* is one of those constructs—the idea that elements of a

complex system organize themselves. There is indeed evidence of self-organization in cognitive activity. For example, the idea of self-organization has been invoked to address the origins of language (e.g., [70]), to observe the emergence of knowledge (e.g., [95]), to explain the systematic problem-solving behaviors of infants (e.g., [96, 97]), and to apply effective pedagogy [98]. Hence, it is reasonable to assume that knowledge is self-organized.

Another complexity construct that anticipates knowledge organization is *self-similarity*—the idea that an organized pattern repeats itself at various nested levels. Here too there is evidence that self-similarity applies to cognition. It was studied primarily by looking for *scale-free patterns* in cognitive behavior [99]. The signature of scale-free pattern is a *1/f scaling*, also known as *pink noise* (e.g., [100, 101]). Analyses of the variability in reaction time have revealed pink-noise patterns, indicating that the variability in a short time series is similar to that in a longer time series (e.g., [102]). Hence, it is reasonable to assume that knowledge is organized in scale-free patterns.

4.1.2 Link 2: Complexity in the qualitative difference in knowledge structures

There is no obvious complexity construct that capture the distinctions between different types of knowledge. At the same time, the complexity angle constrains the

	Nature of Knowledge		**Acquisition of Knowledge**		**Change of Knowledge**	
	Structure	**Diversity**	**Construal**	**Context**	**Persistence**	**Conflict**
Constructs from Non-Living Systems						
Self-organization	**X**		**X**			
Chaos				**X**		
Hysteresis				**X**	**X**	
Attractors					**X**	
Self-organized criticality		**X**		**X**		
Self-similarity	**X**					
Constructs from Living Systems						
Affordance			**X**			
Synchrony			**X**			
Self-preservation					**X**	
Constructs from Thermodynamic Systems						
Balance/ Equilibrium					**X**	**X**
Dissipation pressure						**X**
Autocata kinetics/ Teleodynamics					**X**	

*Note: The columns correspond to the six knowledge truisms described in **Table 1**. The rows correspond to the complexity constructs described in **Table 2**. The X marks the proposed relevance of a complexity construct for a given knowledge truism.*

Table 3.
Relevance of complexity constructs to knowledge formation.

ways in which organizations can differ. For example, given that complex systems consist of elements that interact with each other, differences need to be limited to the elements (e.g., number, type) or the way elements interact (e.g., coupling strength). *Graph theory* can specify the number of connections, thus distinguishing between qualitatively different networks (e.g., small-world networks, scale-free networks). And *ascendency* can capture the coupling strength among elements, thus differentiating systems of various stabilities [80].

Applied to cognition, several complexity measures have been developed to capture coupling strength [103, 104]. These include a child's reasoning during a gear-turning task [105], a child's predictions of the faster sinking object [106], and a child's attempts to balance beams on a fulcrum [107]. Thus, it is reasonable to assume that knowledge structures can differ in the number of mental elements and/or in how the mental elements combine. The organization of preconceptions, for example, might be more restricted than the organization of misconceptions.

4.1.3 Link 3: Complexity in the construal of knowledge

There are several complexity constructs that capture the idea of knowledge construal. *Self-organization* is one of these constructs: It states that the system's organized behavior emerges without a direct linear cause–effect relation. Thus, it rejects the idea that an outside force can specify the exact details of the system's organization. Work on children's stepping behavior has provided early evidence for this conceptualization [108]. More generally, knowledge construal is likely to be self-organized, too.

Affordance is another complexity construct that emphasizes the separation between outside forces and internal organization. This construct rejects the idea altogether that there is objective outside information. Affordances are instead intricately linked to the agent's actions and action capabilities, and thus exist as part of the agent's knowledge structure. In the field of cognition, the concept of affordance can be seen in research of networks that explain decision making, working memory, and mental representations [109–111]. Thus, it is possible that knowledge construal is analogous to the emergence of an affordance.

The construct of *synchrony* hints at a possible mechanism by which a system's organization could be construed. It captures the idea that elements affect each other in a mutually constraining way. This resulting interdependence of elements can amplify the initial coordination to the point that it no longer reflects the outside that gave rise to it (see also *interaction-dominant cognition*; [112, 113]). Synchrony has been used to map out neural connections (see also *connectome*; [114]) and the neural networks that give rise to cognitive performance [115–117]. More generally, there is evidence of synchronization between brain activity and the body/physiology that has been used to capture cognition (e.g., [118, 119]).

4.1.4 Link 4: Complexity in the context dependence of knowledge acquisition

There are several complexity constructs that anticipate context effects (i.e., that seemingly irrelevant changes in context can affect children's learning). Consider, for example, the construct of *self-organized criticality*. This construct describes a system that has several different possible organizations available, which are decided upon by only miniscule changes in the context. Thus, context effects are at the essence of this complexity construct. Indeed, there is evidence that self-organized criticality plays a role in knowledge formation ([113, 120, 121]; see also *metastability*; *multistability*; [122, 123]). Therefore, the context effects seen during learning might be the result of such self-organized criticality.

More generally, the power of seemingly irrelevant aspects of the outside are highlighted by the constructs of *chaos* (i.e., sensitivity to initial conditions) and *hysteresis* (i.e., sensitivity to the history of the system). Here again there is evidence that these concepts are applicable to cognitive processes [124]. Stamovlasis [125], for example, has demonstrated hysteresis in students' science learning, modulated by parameters such as logical thinking ability. Thus, it is possible that context effects seen during learning might be the result of the inherent complexity of knowledge formation.

4.1.5 Link 5: Complexity in the persistence of knowledge structures

There are several complexity constructs that anticipate persistence in the organization of a system's elements. *Hysteresis* is an example of such a construct, namely, because it captures the lingering of a specific organization past outside changes. The construct of *attractors* captures the idea of persistence more generally—that a system's organization can resist perturbation and return to its preferred behavior once the perturbation ends. Applied to children's cognition, the idea of an attractor was used to explain perseverative search behavior [126]. It has also been examined in the study of recurrent neural networks [127, 128]. Thus, it is reasonable to assume that knowledge persistence is the result of an attractor.

The constructs of agency, *autopoiesis, autocatakinetics*, and *teleodynamics* have also been linked to human behavior [129, 130] and mental activity (e.g., [83, 85, 130–134]). In fact, Barab et al. [131] have applied the idea of autocatakinetics specifically to children's science learning.

4.1.6 Link 6: Complexity in the role of conflict in conceptual change

There are two complexity constructs that anticipate the power of conflict to change a system's organization: that of *balance* and *dissipation pressure*. Both of these constructs stem from the study of thermodynamic systems. Under this framework, the perceived conflict can be conceptualized as something that changes the balance of forces and, thus, changes the dissipation pressure. These changes, in turn, affect the likelihood that an existing organization can no longer dissipate the pressure, ushering the change in organization.

The concept of balance is not foreign to work on children's cognition [135]. For instance, Piaget's constructivist account of cognitive disequilibrium highlighted the interplay of the counteracting processes of transformation and conservation [136, 137]. Also, Piaget's notion of adaptation is seen as a process of equilibration between processes of assimilation and accommodation [138]. The role of perceived conflict fits well within this line of work. Thus, the complexity angle offers a way of conceptualizing the role of conflict in ways that are consistent with systemic laws.

4.2 Summary of how complexity is linked to knowledge formation

In this section, we sought to explore the extent to which selected knowledge truisms align with complexity constructs. Our analysis showed that this link is indeed present, though to various degrees: Most prevalently, complexity anticipates the organization of elements and the persistence of knowledge. It also anticipates the influence of the outside context and the impact of conflict on conceptual change. Note, however, that complexity constructs differed in how well they covered knowledge truisms. For example, the idea of knowledge construal was covered by several complexity constructs, while the idea of knowledge persistence was covered primarily by thermodynamic constructs. It remains to be seen if this disparity

identifies a shortcoming of the current theorization of complexity or our interpretation of knowledge findings.

5. Implications for science learning

Having provided an alignment between complexity constructs and knowledge formation, we now derive complexity answers to the ongoing questions related to children's science learning. We specifically focus on questions of (1) how to best define knowledge, (2) how to support children's learning, and (3) how to replace children's mistaken beliefs with scientifically valid insights.

5.1 How to best define knowledge and its elements

While it is widely accepted that knowledge is more than a set of isolated factoids, there is uncertainty about how to best conceptualize such interconnected whole. Complexity provides important constraints for the depiction of knowledge. By this conceptualization, knowledge is defined as the coordination among elements, analogous to a set of synchronizing metronomes, a flock of birds, or an ecosystem. That is to say, knowledge is stable only in the continuous interaction among mental elements. Accordingly, **Figure 1** might need to be revised: Whether understanding is naïve or competent, mutually constraining interactions among elements are required in both. Even elements might be synchronized patterns of interacting parts.

There is also uncertainty about how to capture different types of knowledge unequivocally—for example, between novices and experts. In the balance-beam task, for example, it is still debated whether the difference between implicit and explicit knowledge spans four levels [139], seven levels [140], or none at all [141]. Complexity sheds light on the matter by specifying the ways in which structures can differ. Correspondingly, implicit knowledge might consist of few elements that are constrained to a local action. Explicit knowledge, in contrast, might involve elements that span various circumstances and thus couple with each other on the basis of symbolic correspondences that can be verbalized.

5.2 How to support children's learning

There is no agreed-upon understanding of the processes that turn information into knowledge. Complexity science specifies that this process involves the synchronization of experiences into a self-sustaining whole. Furthermore, thermodynamic constructs show that such synchronized aggregations emerge when there is a balance between clustered energy and pressure. Thus, to decide on the ideal pedagogy, one must first identify the 'clustered energy' in the learning context, as well as the nature of 'pressure'. One must then ensure that these two aspects are in some sort of equilibrium to allow for learning.

Applied to the balance-beam task, clustered energy could be conceptualized as information about the beams (visual, haptic). There is also information across trials, for example, that some of the beams balance at their geometric center. The pressure, on the other hand, could be conceptualized as the task that children are asked to complete: to balance individual beams on a fulcrum. The narrower the fulcrum, the more pressure there is on the system to organize its elements. For pedagogy to be effective, therefore, the salience of the beam's weight distribution must be calibrated with the narrowness of the fulcrum upon which the beam should be balanced. This calibration between information and task pressure has to fit the competence of the individual child and adjust flexibly to changing competences.

5.3 How to replace mistaken beliefs with scientifically valid insights

The challenge in science education has been largely attributed to the presence of mistaken beliefs. However, the results are mixed on the recommendation to assess existing beliefs first, prior to administering a science lesson [142–144]. Complexity science can again shed light on the issue. Specifically, lessons derived from thermodynamics provide a cautionary note to the logic of first providing children with an assessment. This is because, in the language of complexity, assessments are equivalent to the pressure on the system to organize itself. This pressure might force children to come up with ordered behavior that resembles a belief. The risk, therefore, is that the assessment pushes the learner to form an ad-hoc belief, rather than assessing the presence of an already existing belief.

The solution lies in combining pressure (the assessment) with support (the information relevant to the solution), rather than offering the assessment on its own. This recommendation is in line with the resubsumption theory [144, 145]. It is also in line with the finding that a child's explicit goal to change mistaken beliefs has a positive effect on learning [146–148]. This is because such explicit buy-in from the learner shifts the nature of the pressure in ways that allows children to actively search for scientifically valid patterns (vs. latch onto the most obvious patterns to coordinate experiences).

Ultimately, the complexity viewpoint implies that the challenge of science learning lies in the nature of science itself, rather than in the presence of mistaken beliefs. This is because the patterns of order relevant to science concepts are often hidden behind more salient but irrelevant science concepts. For example, in the case of balance beams, visual features are likely to have priority over haptic features, making the irrelevant aspect of the beam's shape more readily available than the relevant weight distribution. Therefore, to improve science learning, one would need to invest in ways of making relevant patterns of order more salient than irrelevant ones, paired with gearing children's action toward detecting these relevant patterns.

5.4 Summary of complexity-based answers to open questions

In this section, we sought to address practical implications of a complexity view of learning. On the question of the nature of knowledge, for example, complexity science provides details on how to conceptualize the interaction of mental elements that gives rise to knowledge. And on the question of learning, complexity science can pin down the pedagogy that could help children ignore irrelevant aspects of the context. The complexity angle can even address questions about conceptual change: It undermines the common suggestion of assessing children's naïve beliefs in the absence of instruction; and it highlights strategies that can help children learn about abstract science concepts. While these suggestions are merely hinted at, they can offer an important impetus to science-education research.

6. Conclusion

In line with the volume's goal of deepening the meaning of complexity, we traced the connection between complexity constructs and children's learning. Our specific focus was on children's science education, a topic with remaining open questions despite previous attempts to apply complexity ideas. Our rationale was that neither the field of complexity nor the field of children's learning are streamlined: Both areas feature inconsistencies and gaps [149]. The synthesis we offered was designed to substantiate this link, potentially fostering progress in both fields.

Our approach started with a preliminary step—namely, to consider issues of cognition separately from issues of complexity. To this end, we defined central characteristics of knowledge formation without considerations of complexity; and we defined central characteristics of complex systems without considerations of cognition. This two-pronged preliminary step made it possible to explore the link between complexity and learning in a principled way, rather than trying to prove a-priori assumptions about it. Thus, by cross-tabulating the list of knowledge truisms with the list of complexity constructs, we were able to substantiate the knowledge–complexity link in a relatively objective way.

The cross-tabulation shows that our chosen knowledge truisms were anticipated robustly by complexity constructs. Building on this alignment, we were able to derive answers relevant to science education. For example, the knowledge-complexity alignment specifies that knowledge is a mental synchronization of experiences. Such synchronization can emerge when there is a balance between direct instruction and active learning that is calibrated to highlight relevant patterns of order (vs. irrelevant patterns of order). This calibration can be difficult to establish when relevant patterns are inherently hidden, as is the case in abstract science concepts. In turn, this difficulty can explain the challenge of science education, going against the prevailing assumption that science-education challenges stem from children's misconceptions.

A limitation of this work pertains to taking some shortcuts when generating the two initial lists. For example, we settled on six knowledge truisms, potentially at the expense of important nuances. And we prioritized prominent complexity constructs, potentially at the expense of lesser-known constructs. We also overlooked ongoing controversies, for example on the topic of constructivism, on self-organized criticality, or on how to apply thermodynamics to cognitive processes. For these reasons, our lists are undoubtedly incomplete. Nevertheless, this work offers a starting point from which to develop a complexity-based framework for children's learning.

Author details

Michael J. Droboniku*, Heidi Kloos, Dieter Vanderelst and Blair Eberhart
Department of Psychology, University of Cincinnati, Cincinnati, OH, United States

*Address all correspondence to: drobonmj@mail.uc.edu

References

[1] diSessa, A. A. (2002). Why "conceptual ecology" is a good idea. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 28-60). Springer. https://doi.org/10.1007/0-306-47637-1_2

[2] Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning.* Harvard University Press.

[3] Carey, B. (2015). *How we learn: The surprising truth about when, where, and why it happens.* Random House Trade Paperbacks.

[4] Harris, P. L. (2000). *The work of the imagination*. Blackwell Publishing.

[5] National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition.* National Academies Press.

[6] Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. Cognition, *3*(3), 195-212. https://doi.org/10.1016/0010-0277(74)90008-0

[7] Larsson, Å., & Halldén, O. (2010). A structural view on the emergence of a conception: Conceptual change as radical reconstruction of contexts. Science Education, *94*(4), 640-664. https://doi.org/10.1002/sce.20377

[8] Sheets-Johnstone, M. (1999). Emotion and movement: A beginning empirical-phenomenological analysis of their relationship. Journal of Consciousness Studies, *6*(12), 259-277.

[9] Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

[10] Craik, K. J. W. (1952). *The nature of explanation* (Vol. 445). CUP Archive.

[11] Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction*. Lawrence Erlbaum Associates, Inc.

[12] Piaget, J., & Inhelder, B. (1974). *The child's construction of quantity*. Routledge & Kegan Paul.

[13] Kalish, M. L., Lewandowsky, S., & Davies, M. (2005). Error-driven knowledge restructuring in categorization. Journal of Experimental Psychology: Learning, Memory, & Cognition, *31*(5), 846-861. https://doi.org/10.1037/0278-7393.31.5.846

[14] Bullock, M. (1985). Causal reasoning and developmental change over the preschool years. Human Development, *28*(4), 169-191. https://doi.org/10.1159/000272959

[15] Rebich, S., & Gautier, C. (2005). Concept mapping to reveal prior knowledge and conceptual change in a mock summit course on global climate change. Journal of Geoscience Education, *53*(4), 355-365. https://doi.org/10.5408/1089-9995-53.4.355

[16] Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. Behavioral & Brain Sciences, *22*(5), 735-808.

[17] Krist, H., Fieberg, E. L., & Wilkening, F. (1993). Intuitive physics in action and judgment: The development of knowledge about projectile motion. Journal of Experimental Psychology: Learning, Memory, & Cognition, *19*(4), 952-966. https://doi.org/10.1037/0278-7393.19.4.952

[18] Bennet, D., & Bennet, A. (2008). The depth of knowledge: Surface, shallow or deep? Vine, *38*(4). 405-420. https://doi.org/10.1108/03055720810917679

[19] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2). MIT Press.

[20] Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. Giere (Ed.), *Cognitive models of science: Minnesota studies in the philosophy of science* (pp. 129-186). University of Minnesota Press.

[21] Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. Journal of Experimental Child Psychology, *49*(1), 31-57. https://doi.org/10.1016/0022-0965(90)90048-D

[22] Pine, K. J., & Messer, D. J. (2000). The effect of explaining another's actions on children's implicit theories of balance. Cognition & Instruction, *18*(1), 35-51. https://doi.org/10.1207/S1532690XCI1801_02

[23] Lang, C., Siemens, G., Wise, A., & Gasevic, D. (Eds.). (2017). *Handbook of learning analytics*. SOLAR - Society for Learning Analytics & Research.

[24] Sawyer, R. K. (Ed.). (2014). *The Cambridge Handbook of the Learning Sciences.* Cambridge University Press.

[25] Bruya, B. (Ed.). (2010). *Effortless attention: A new perspective in the cognitive science of attention and action*. MIT Press.

[26] Folger, R., & Stein, C. (2017). Abduction 101: Reasoning processes to aid discovery. Human Resource Management Review, *27*(2), 306-315. https://doi.org/10.1016/j.hrmr.2016.08.007

[27] Rose, S. A., & Blank, M. (1974). The potency of context in children's cognition: An illustration through conservation. Child Development, *45*(2), 499-502. https://doi.org/10.2307/1127977

[28] Doolittle, P. E. (2014). Complex constructivism: A theoretical model of complexity and cognition. International Journal of Teaching & Learning in Higher Education, *26*(3), 485-498.

[29] Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. Review of Educational Research, *63*(2), 167-199. https://doi.org/10.3102/00346543063002167

[30] Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. International Journal of Science Education, *25*(6), 671-688. https://doi.org/10.1080/09500690305016

[31] Limón, M., & Mason, L. (Eds.). (2002). *Reconsidering conceptual change: Issues in theory and practice*. Springer. https://doi.org/10.1007/0-306-47637-1

[32] Duit, R. (1991). Students' conceptual frameworks: Consequences for learning science. The Psychology of Learning Science, *75*(6), 649-672.

[33] Kummer, T. A., Whipple, C. J., & Jensen, J. L. (2016). Prevalence and persistence of misconceptions in tree thinking. Journal of Microbiology & Biology Education, *17*(3), 389-398. https://doi.org/10.1128/jmbe.v17i3.1156

[34] Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. Science Education, *66*(2), 211-227.

[35] Vosniadou, S. (Ed.). (2009). *International handbook of research on conceptual change*. Routledge.

[36] Piaget, J. (1964). Cognitive development in children: Piaget. Journal of research in science teaching, *2*(3), 176-186.

[37] Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.

[38] Dewey, J. (1997). *How we think*. Courier Corporation.

[39] D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. Learning & Instruction, *29*(1), 153-170. https://doi.org/10.1016/j.learninstruc.2012.05.003

[40] Camazine, S., Deneubourg, J. L., Franks, N. R., Sneyd, J., Bonabeau, E., & Theraula, G. (2003). *Self-organization in biological systems*. Princeton University Press.

[41] Holland, J. H. (2014). *Complexity: A very short introduction*. OUP Oxford.

[42] Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

[43] Newman, M. E. J. (2011). Resource letter CS–1: Complex systems. American Journal of Physics, *79*(8), 800-810. https://doi.org/10.1119/1.3590372

[44] Ziemelis, K. (2001). Complex systems. Nature, *410*(1), 241. https://doi.org/10.1038/35065672

[45] Wolfram, S. (1984). Cellular automata as models of complexity. Nature, *311*(5985), 419-424. https://doi.org/10.1038/311419a0

[46] Tanaka, D. (2007). General chemotactic model of oscillators. Physical Review Letters, *99*(13), 134103. https://doi.org/10.1103/PhysRevLett.99.134103

[47] Bao, Z., Ye, Y., & Wu, L. (2020). Multi-timescale coordinated schedule of interdependent electricity-natural gas systems considering electricity grid steady-state and gas network dynamics. International Journal of Electrical Power & Energy Systems, *118*(1), 105763. https://doi.org/10.1016/j.ijepes.2019.105763

[48] Haken, H. (2008). Self-organization. Scholarpedia, *3*(8), 1401. https://doi.org/10.4249/scholarpedia.1401

[49] Bonabeau, E., Theraulaz, G., Deneubourg, J. L., Aron, S., & Camazine, S. (1997). Self-organization in social insects. Trends in Ecology & Evolution, *12*(5), 188-193. https://doi.org/10.1016/S0169-5347(97)01048-3

[50] Ball, P. (2009). In retrospect: the physics of sand dunes. Nature, *457*(7233), 1084-1085. https://doi.org/10.1038/4571084a

[51] Pantaleone, J. (2002). Synchronization of metronomes. American Journal of Physics, *70*(10), 992-1000. https://doi.org/10.1119/1.1501118

[52] Lorenz, E. (2000). The butterfly effect. World Scientific Series on Nonlinear Science - Series A, *39*(1), 91-94.

[53] Vernon, J. L. (2017). Understanding the butterfly effect. American Scientist, *105*(3), 130.

[54] Levien, R. B., & Tan, S. M. (1993). Double pendulum: An experiment in chaos. American Journal of Physics, *61*(11), 1038-1044. https://doi.org/10.1119/1.17335

[55] Mellodge, P. (2016). *A practical approach to dynamical systems for engineers*. Woodhead Publishing. https://doi.org/10.1016/C2014-0-03574-5

[56] Poston, T., & Stewart, I. (2014). *Catastrophe theory and its applications*. Courier Corporation.

[57] Sharov, A. (1996). Attractors and their types. Retrieved from https://web.ma.utexas.edu/users/davis/375/popecol/lec9/attract.html

[58] Bak, P., Tang, C., & Wiesenfeld, K. (1988). Self-organized criticality. Physical Review A, *38*(1), 364. https://doi.org/10.1103/PhysRevA.38.364

[59] Bak, P., & Chen, K. (1991). Self-organized criticality. Scientific American, *264*(1), 46-53. https://www.jstor.org/stable/24936753

[60] Mandelbrot, B. B. (1982). *The fractal geometry of nature* (Vol. 1). New York: WH Freeman.

[61] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. Science, *286*(5439), 509-512. https://doi.org/10.1126/science.286.5439.509

[62] de Solla Price, D. (1976). A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science, *27*(5), 292-306. https://doi.org/10.1002/asi.4630270505

[63] Davies, J. A. (2014). *Life unfolding: How the human body creates itself*. OUP Oxford.

[64] Goodwin, B. (2020). *How the leopard changed its spots: The evolution of complexity*. Princeton University Press.

[65] Gibson, J. J., & Carmichael, L. (1966). *The senses considered as perceptual systems* (Vol. 2, No. 1, pp. 44-73). Houghton Mifflin.

[66] Srinivasan, M., Zhang, S., Lehrer, M., & Collett, T. S. (1996). Honeybee navigation en route to the goal: Visual flight control and odometry. Journal of Experimental Biology, *199*(1), 237-244. https://jeb.biologists.org/content/199/1/237

[67] Ravignani, A. (2017). Agree on definitions of synchrony. Nature, *545*(1), 158. https://doi.org/10.1038/545158c

[68] Strogatz, S. H. (2012). *Sync: How order emerges from chaos in the universe, nature, and daily life.* Hachette UK.

[69] Darwin, C. (1859). *On the origins of species by means of natural selection: Or, the preservation of favoured races in the struggle for life*. John Murray.

[70] Steels, L. (2015). *The talking heads experiment: Origins of words and meanings* (Vol. 1). Language Science Press.

[71] Maturana, H. R., & Varela, F. J. (1991). *Autopoiesis and cognition: The realization of the living* (Vol. 42). Springer Science & Business Media.

[72] Bickhard, M. H. (2000). Autonomy, function, and representation. Communication & Cognition – Artificial Intelligence, *17*(4), 111-131.

[73] Ulanowicz, R. E. (1997). Ecology, the ascendent perspective. Columbia University Press.

[74] Kondepudi, D. (2012). Self-organization, entropy production, and physical intelligence. Ecological Psychology, *24*(1), 33-45. https://doi.org/10.1080/10407413.2012.643716

[75] Kondepudi, D., Kay, B., & Dixon, J. (2015). End-directed evolution and the emergence of energy-seeking behavior in a complex system. Physical Review E, *91*(5), 050902. https://doi.org/10.1103/PhysRevE.91.050902

[76] Müller, I. (2007). *A history of thermodynamics: The doctrine of energy and entropy*. Springer Science & Business Media.

[77] Jørgensen, S. E. (2000). *Thermodynamics and ecological modelling*. CRC press.

[78] Rennie, R., & Law, J. (Eds.). (2019). *A dictionary of physics* (8th ed.). Oxford University Press. https://doi.org/10.1093/acref/9780198821472.001.0001

[79] Brinck, K. & Jenson, H. J. (2017). The evolution of ecosystem ascendency in a complex systems based model. Journal of Theoretical Biology, *428*(1), 18-25. https://doi.org/10.1016/j.jtbi.2017.06.010

[80] Ulanowicz, R. E. (2000). Ascendancy: A measure of ecosystem performance. In S. E. Jorgensen (Ed). *Handbook of ecosystem theories and management* (pp 303-316). CRC press.

[81] Ulanowicz, R. E. (2012). *Growth and development: Ecosystems phenomenology*. Springer Science & Business Media.

[82] Schneider, E. D., & Kay, J. J. (1994). Complexity and thermodynamics: Towards a new ecology. Futures, *26*(6), 626-647. https://doi.org/10.1016/0016-3287(94)90034-5

[83] Deacon, T. W. (2011). *Incomplete nature: How mind emerged from matter*. WW Norton & Company.

[84] Taleb, N. N. (2012). *Antifragile: Things that gain from disorder* (Vol. 3). Random House Incorporated.

[85] Swenson, R. (1997). Autocatakinetics, evolution, and the law of maximum entropy production: A principled foundation towards the study of human ecology. Advances in Human Ecology, *6*(1), 1-48.

[86] Sagan, D. (2012). Teleodynamics. Tartu Semiotics Library, *1*(10), 291-295.

[87] Drazin, R., & Sandelands, L. (1992). Autogenesis: A perspective on the process of organizing. Organization Science, *3*(2), 230-249. https://doi.org/10.1287/orsc.3.2.230

[88] Eigen, M., & Schuster, P. (2012). *The hypercycle: A principle of natural self-organization*. Springer Science & Business Media.

[89] Morin, E. (1992). From the concept of system to the paradigm of complexity. Journal of Social & Evolutionary Systems, *15*(4), 371-385. https://doi.org/10.1016/1061-7361(92)90024-8

[90] Davis, B., & Sumara, D. J. (2006). *Complexity and education: Inquiries into learning, teaching, and research*. Psychology Press.

[91] Elman, J. L., Bates, E. A., & Johnson, M. H. (1996). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). MIT Press.

[92] Niebaum, J., & Munakata, Y. (2020). Deciding what to do: Developments in children's spontaneous monitoring of cognitive demands. Child Development Perspectives, *14*(4), 202-207. https://doi.org/10.1111/cdep.12383

[93] Van Geert, P., & Steenbeek, H. (2005). Explaining after by before: Basic aspects of a dynamic systems approach to the study of development. Developmental Review, *25*(4), 408-442. https://doi.org/10.1016/j.dr.2005.10.003

[94] Wilensky, U., & Jacobson, M. J. (2014). Complex systems and the learning sciences. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences, Second Edition* (pp. 319-338). Cambridge University Press.

[95] Dixon, J. A., Stephen, D. G., Boncoddo, R., & Anastas, J. (2010). The self-organization of cognitive structure. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 52, pp. 343-384). Academic Press. https://doi.org/10.1016/S0079-7421(10)52009-7

[96] Corbetta, D., & Thelen, E. (1996). The developmental origins of bimanual coordination: A dynamic perspective. Journal of Experimental Psychology: Human Perception & Performance, *22*(2), 502-522. https://doi.org/10.1037/0096-1523.22.2.502

[97] Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. Behavioral & Brain Sciences, *24*(1), 1-34.

[98] Koopmans, M. (2014). Nonlinear change and the black box problem in educational research. Nonlinear Dynamics, Psychology and Life Sciences, *18*(1), 5-22.

[99] Stadnitski, T. (2012). Measuring fractality. Frontiers in Physiology, *3*(1), 127. https://doi.org/10.3389/fphys.2012.00127

[100] Farrell, S., Wagenmakers, E. J., & Ratcliff, R. (2006). 1/*f* noise in human cognition: Is it ubiquitous, and what does it mean? Psychonomic Bulletin & Review, *13*(4), 737-741. https://doi.org/10.3758/BF03193989

[101] Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/*f* noise in human cognition. Science, *267*(5205), 1837-1839. https://doi.org/10.1126/science.7892611

[102] Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. Journal of Experimental Psychology: General, *132*(3), 331-350. https://doi.org/10.1037/0096-3445.132.3.331

[103] Cox, R. F., van der Steen, S., Guevara, M., de Jonge-Hoekstra, L., & van Dijk, M. (2016). Chromatic and anisotropic cross-recurrence quantification analysis of interpersonal behavior. In *Recurrence plots and their quantifications: expanding horizons* (pp. 209-225). Springer, Cham. https://doi.org/10.1007/978-3-319-29922-8_11

[104] Jonge-Hoekstra, D., Van der Steen, S., Van Geert, P., & Cox, R. F. (2016). Asymmetric dynamic attunement of speech and gestures in the construction of children's understanding. Frontiers in Psychology, *7*(1), 473. https://doi.org/10.3389/fpsyg.2016.00473

[105] Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: Entropy, action, and cognition. Journal of Experimental Psychology: Human Perception & Performance, *35*(6), 1811-1832. https://doi.org/10.1037/a0014510

[106] Castillo, R. D., Kloos, H., Richardson, M. J., & Waltzer, T. (2015). Beliefs as self-sustaining networks: Drawing parallels between networks of ecosystems and adults' predictions. Frontiers in Psychology, *6*(1), 1723. https://doi.org/10.3389/fpsyg.2015.01723

[107] Fleuchaus, E., Kloos, H., Kiefer, A. W., & Silva, P. L. (2020). Complexity in science learning: Measuring the underlying dynamics of persistent mistakes. The Journal of Experimental Education, *88*(3), 448-469. https://doi.org/10.1080/00220973.2019.1660603

[108] Thelen, E., Smith, L. B., Lewkowicz, D. J., & Lickliter, R. (1994). *A dynamic systems approach to the development of cognition and action* (Vol. 10). Cambridge, MA: MIT press.

[109] Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K., ... & Zeschel, A. (2010). Integration of action and language knowledge: A roadmap for developmental robotics. IEEE Transactions on Autonomous Mental Development, *2*(3), 167-195. https://doi.org/10.1109/TAMD.2010.2053034

[110] Pezzulo, G., & Cisek, P. (2016). Navigating the affordance landscape: feedback control as a process model of behavior and cognition. Trends in Cognitive Sciences, *20*(6), 414-424. http://dx.doi.org/10.1016/j.tics.2016.03.013

[111] Yamashita, Y., & Tani, J. (2008). Emergence of functional hierarchy in a multiple timescale neural network model: A humanoid robot experiment. PLoS Comput Biol, *4*(11), e1000220. https://doi.org/10.1371/journal.pcbi.1000220

[112] Richardson, M. J., & Chemero, A. (2014). Complex dynamical systems and embodiment. In L. Shapiro (Ed.). *The Routledge handbook of embodied cognition* (pp. 39-50). Routledge Taylor & Francis Group.

[113] Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2005). Human cognition and 1/*f* scaling. Journal of Experimental Psychology: General, *134*(1), 117-123. https://doi.org/10.1037/0096-3445.134.1.117

[114] Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: a structural description of the human brain. PLoS Comput Biol, *1*(4), e42. https://doi.org/10.1371/journal.pcbi.0010042

[115] Alderson, T. H., Bokde, A. L., Kelso, J. S., Maguire, L., & Coyle, D. (2020). Metastable neural dynamics underlies cognitive performance across multiple behavioural paradigms. Human Brain Mapping, *41*(12), 3212-3234. https://doi.org/10.1002/hbm.25009

[116] Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. Neuroimage, *52*(3), 1059-1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

[117] Schoner, G., & Kelso, J. A. (1988). Dynamic pattern generation in behavioral and neural systems. Science, *239*(4847), 1513-1520. https://doi.org/10.1126/science.3281253

[118] Byrge, L., Sporns, O., & Smith, L. B. (2014). Developmental process emerges from extended brain–body–behavior networks. Trends in Cognitive Sciences, *18*(8), 395-403. https://doi.org/10.1016/j.tics.2014.04.010

[119] Smith, L., Byrge, L., & Sporns, O. (2020). Beyond origins: Developmental pathways and the dynamics of brain networks. In A. J. Lerner, S. Cullen, & S. J. Leslie (Eds.), *Current controversies in philosophy of cognitive science* (pp. 49-62). Routledge.

[120] Hellyer, P. J., Scott, G., Shanahan, M., Sharp, D. J., & Leech, R. (2015). Cognitive flexibility through metastable neural dynamics is disrupted by damage to the structural connectome. Journal of Neuroscience, *35*(24), 9050-9063. https://doi.org/10.1523/JNEUROSCI.4648-14.2015

[121] Tadić, B., Dankulov, M. M., & Melnik, R. (2017). Mechanisms of self-organized criticality in social processes of knowledge creation. Physical Review E, *96*(3), 032307. https://doi.org/10.1103/PhysRevE.96.032307

[122] Feudel, U., & Grebogi, C. (1997). Multistability and the control of complexity. Chaos: An Interdisciplinary Journal of Nonlinear Science, *7*(4), 597-604. https://doi.org/10.1063/1.166259

[123] Kelso, J. S. (2012). Multistability and metastability: Understanding dynamic coordination in the brain. Philosophical Transactions of the Royal Society B: Biological Sciences, *367*(1591), 906-918. https://doi.org/10.1098/rstb.2011.0351

[124] Tognoli, E., & Kelso, J. S. (2014). The metastable brain. Neuron, *81*(1), 35-48. https://doi.org/10.1016/j.neuron.2013.12.022

[125] Stamovlasis, D. (2014). Bifurcation and hysteresis effects in student performance: The signature of complexity and chaos in educational research. Complicity: An International Journal of Complexity & Education, *11*(2). https://doi.org/10.29173/cmplct22964

[126] Spencer, J. P., Smith, L. B., & Thelen, E. (2001). Tests of a dynamic systems account of the A-not-B error: The influence of prior experience on the spatial memory abilities of two-year-olds. Child Development, *72*(5), 1327-1346. https://doi.org/10.1111/1467-8624.00351

[127] Cabessa, J., & Villa, A. E. (2014). An attractor-based complexity measurement for boolean recurrent neural networks. PLoS One, *9*(4), e94204. https://doi.org/10.1371/journal.pone.0094204

[128] Cabessa, J., & Villa, A. E. (2016). Expressive power of first-order recurrent neural networks determined by their attractor dynamics. Journal of Computer & System Sciences, *82*(8), 1232-1250. https://doi.org/10.1016/j.jcss.2016.04.006

[129] Juarrero, A. (2000). Dynamics in action: Intentional behavior as a complex system. Emergence, *2*(2), 24-57. https://doi.org/10.1207/S15327000EM0202_03

[130] Kelso, J. S. (2016). On the self-organizing origins of agency. Trends in Cognitive Sciences, *20*(7), 490-499. https://doi.org/10.1016/j.tics.2016.04.004

[131] Barab, S. A., Cherkes-Julkowski, M., Swenson, R., Garrett, S., Shaw, R. E., & Young, M. (1999). Principles of self-organization: Learning as participation in autocatakinetic systems. Journal of the Learning Sciences, *8*(4), 349-390. https://doi.org/10.1080/10508406.1999.9672074

[132] Swenson, R., & Turvey, M. T. (1991). Thermodynamic reasons for perception–action cycles. Ecological Psychology, *3*(4), 317-348. https://doi.org/10.1207/s15326969eco0304_2

[133] Turvey, M. T., & Carello, C. (2012). On intelligence from first principles: Guidelines for inquiry into the hypothesis of physical intelligence (PI). Ecological Psychology, *24*(1), 3-32. https://doi.org/10.1080/10407413.2012.645757

[134] Ulanowicz, R. E. (2020). Ecological clues to the nature of consciousness. Entropy, *22*(6), 611. https://doi.org/10.3390/e22060611

[135] Appel, M. H., & Goldberg, L. S. (1977). *Topics in cognitive development. (vol 1) - equilibration: Theory, research, and application.* Plenum Press.

[136] Martin, K., Simpson, D. J., & Gallagher, J. (1998). Piaget, Dewey, and Complexity. Journal of Thought, *33*(2), 75-82.

[137] Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures.* (Trans A. Rosin). Viking.

[138] DiPaolo, E., Buhrmann, T., & Barandiaran, X. (2017). *Sensorimotor life: An enactive proposal.* Oxford University Press.

[139] Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.

[140] Pine, K., & Messer, D. (2003). The development of representations as

children learn about balancing. British Journal of Developmental Psychology, *21*(2), 285-301. https://doi.org/10.1348/026151003765264093

[141] Krist, H., Horz, H., & Schönfeld, T. (2005). Children's block balancing revisited: No evidence for representational redescription. Swiss Journal of Psychology, *64*(3), 183-193. https://doi.org/10.1024/1421-0185.64.3.183

[142] Guzzetti, B. J., & Hynd, C. R. (Eds.). (2013). *Perspectives on conceptual change: Multiple ways to understand knowing and learning in a complex world*. Routledge.

[143] Hewson, P. W., & Hewson, M. G. B. (1984). The role of conceptual conflict in conceptual change and the design of science instruction. Instructional Science, *13*(1), 1-13.

[144] Ohlsson, S. (2011). *Deep learning: How the mind overrides experience*. Cambridge University Press.

[145] Ohlsson, S. (2009). Resubsumption: A possible mechanism for conceptual change and belief revision. Educational Psychologist, *44*(1), 20-40. https://doi.org/10.1080/00461520802616267

[146] Gunstone, R. F., & Mitchell, I. J. (2005). Metacognition and conceptual change. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Teaching science for understanding: A human constructivist view* (pp. 133-163). Academic Press. https://doi.org/10.1016/B978-012498360-1/50006-4

[147] Mason, L. (2001). Introducing talk and writing for conceptual change: A classroom study. Learning & Instruction, *11*(5), 305-329. https://doi.org/10.1016/S0959-4752(00)00035-9

[148] Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. Journal of Educational Psychology, *92*(3), 544-555. https://doi.org/10.1037/0022-0663.92.3.544

[149] Witherington, D. C., & Margett, T. E. (2011). How conceptually unified is the dynamic systems approach to the study of psychological development? Child Development Perspectives, *5*(4), 286-290. https://doi.org/10.1111/j.1750-8606.2011.00211.x

Section 2

Theory of Complexity in Natural Systems

Chapter 4

Metrics Based on Information Entropy to Evaluate Landscape Complexities

Sérgio Henrique Vannucchi Leme de Mattos, Luiz Eduardo Vicente, Andrea Koga Vicente, Cláudio Bielenki Junior, Maristella Cruz de Moraes, Gabriele Luiza Cordeiro and José Roberto Castilho Piqueira

Abstract

Information entropy concept is the base for many measures used to evaluate the complexity of complex environmental systems. Its application has great potential to evaluate landscape organization and dynamics, especially if we consider that there is a direct relation between their patterns and processes: the spatial arrangement (structure) of units within a mosaic reflects on system functions. Consequently, changes on structure reflects on functions and *vice versa*. Here, we exemplify how three measures based on information entropy – LMC and SDL complexity measures and H_e/H_{max} variability measure – could be applied to evaluating the degree of complexity of a landscape and its components by associating their heterogeneity with the diversity of information acquired from the remote sensors' images. For this, we developed two scripts for a Geographical Information System (QGIS): (1) *CompPlex H_eROI*, that compares the complexity of a landscape patch with others and also with their transition areas; and (2) *CompPlex Janus*, which analyzes how complexity varies in the landscape over space and time, generating landscape complexity maps. We also use LMC and SDL complexity measures and H_e/H_{max} variability measure to evaluate complexity time series of environmental variables, as rain and temperature, which allow to evaluate how their variations along time and space affects landscape dynamics. Therefore, application of such metrics in multitemporal studies of landscape dynamics provides indicators of landscape resilience and the degree of conservation or degradation of its different fragments due to anthropic impacts related to land uses.

Keywords: complexity, Information entropy, landscape metrics

1. Introduction

From the perspective of the Complexity Paradigm [1], the landscape can be interpreted as a complex environmental system that is established from the interdependence relationships of the physical-natural system (that is, by the elements and processes present in nature) and the socioeconomic system (that is, the

elements and processes linked to human societies in their cultural, economic and social aspects). When these two systems interact, they are considered as subsystems of a system with a higher level of ecological organization: the landscape. Landscape can be considered the 2nd level of ecological organization, characterized by a set of interrelated ecosystems and formed, as the other levels, by the interactions between society and nature (**Figure 1**).

As is typical of complex adaptive system, the landscape presents non-linear negative and positive feedback processes generated from self-organization of its elements in interaction networks. The structure and dynamics of the landscape are affected and affect the other three levels (holons) constituents of this holarchic organization through bottom-up and top-down processes (**Figure 1**). The evidence of these interactions can be seen in landscape's patterns, as in this type of system, there is a direct relationship between its patterns and processes [4]: the spatial arrangement (structure) of units within a mosaic influence system functions. Consequently, changes in structure reflects on functions and vice versa, therefore affecting landscape resilience and integrity.

Thus, the complexity of a landscape and of the units that comprise the mosaic can be associated with the heterogeneities of its spatial, temporal, and structural patterns, with greater complexities being represented by patterns located in regions of intermediate heterogeneity in a gradient that goes from totally ordered patterns up to those completely disordered [5, 6]. To capture this typical signature of complex environmental systems not only qualitatively, it is necessary to use indicators capable of representing it in a quantitative way. This can be done through measures based on information entropy, which can be applied to assess the structure and dynamics of landscapes, as done by Mattos et al. [7] and Piqueira et al. [8] in

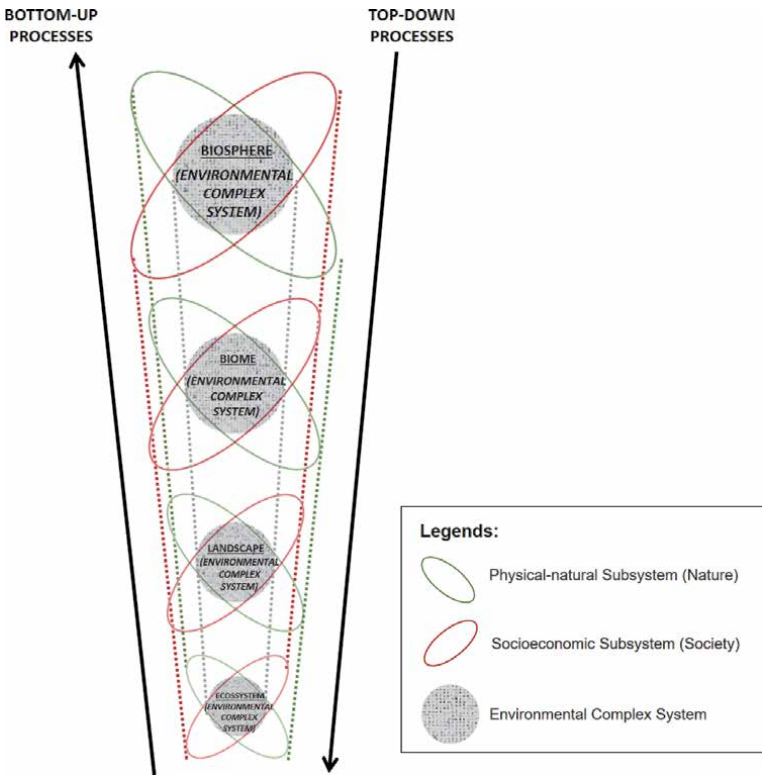


Figure 1.
Holarchic organization levels of complex environmental systems (inpired in: [2, 3].

relation to ecological interactions between their populations or by Piqueira & Mattos [9] for abiotic factors present in them.

Here, the application of measures based on information entropy for two purposes is demonstrated. The first is to assess the complexity of climatic time series. The other is both in comparing the complexity of a landscape patch with others and also with their transition areas, as well as allowing to verify as complexity varies in the landscape over space and time.

2. Use of measures based on information entropy to evaluate landscapes complexity

Remote sensor images can be used to identify landscape patterns in two different ways. The first one is related to the degree of roughness caused by the variability of tone or color, which provides image geometry and texture from targets [10–12]. The other is by examining the image spectral features, analyzing the values (e.g., surface reflectance, radiance, digital numbers) for each wavelength interval (i.e., band) vs. matrix pixels obtained for a specific target, which provides the spectral signature of a specific target [11, 13].

Several methods have been developed for both texture and spectral analysis to recognize patterns in remote sensing data [14–16]. Generally, many measures (also called metrics) derived from these methods have the same purposes: identifying similar patterns that occur in different places and distinguishing different patterns within a landscape. Although theoretically simple, in practice, this objective is not always easily achieved. Here, we discuss the use of three metrics derived from information entropy to measure the complexity of landscape patterns and show their applications to some case studies.

2.1 Information entropy applied to landscape patterns recognition and the evaluation of their level of complexity

Several metrics applied to textural and spectral analysis try to capture landscape patterns by using approaches deriving from theories and methods associated to the complexity paradigm, such as General System Theory, Cybernetic, Theory of Dissipative Structures, Hierarchy Theory, Percolation Theory, Self-Organized Criticality, Catastrophe Theory and Fractal Geometry [17–20].

Information entropy and other measures derived from it are also extensively used to quantify landscape heterogeneity and, consequently, to evaluate its organization level and complexity [18, 21]. According to Shiner et al. [22], there are three broad categories in which complexity measures based on information entropy may be classified: the first is composed of measures that consider complexity as a direct function of disorder (as is the case of very popular Shannon diversity index). So, measures of this category attribute lower values of complexity to ordered states and higher values to disordered states [22, 23]. Another category inverts this interpretation by associating higher complexity to most ordered states [22].

However, both measure categories are considered inadequate since there is no real complexity in situations that present zero or maximum entropy [23]. This fact is particularly applicable in Landscape Ecology, since, as mentioned by Parrot [24], more spatially complex landscapes are those in which the spatial pattern is situated in regions of intermediary heterogeneity, between order and disorder patterns. Thus, the maximum complexity would be located between these two extreme situations, which could be mathematically expressed as a convex function of disorder, as are the measures belonging to the third category defined by Shiner et al. [22].

This is the case of the LMC and SDL measures, proposed initially by López-Ruiz et al. [25] and Shiner et al. [22], respectively. In both definitions, the two complementary parameters – disorder and order – are combined to obtain a complexity measure.

2.1.1 H_e/H_{max}, LMC and SDL complexity measures and their application to remote sensing images

Shannon's Information Theory [26] could be applied to reflectance data in a remote sensing image band. These data are represented by their discretization in single digital numbers (DN), each DN representing a pixel value related to the intensity of the radiation in a particular wavelength at the sensor [11]. As the occurrence of certain DN values becomes more likely than other values, the entropy of the image decreases.

The variability measure H_e/H_{max} is related to Shannon entropy, and it belongs to the first category mentioned by Shiner et al. [22]; therefore, considering that complexity increases as a function of increasing the system disorder. This measure is useful to verify if a landscape and its patches are near the ordered/homogeneous or the disordered/heterogeneous patterns. To use this measure, it is necessary first to define system extension (N), given by the system's total number of possible states. In the case of remote sensing images, N corresponds to the number of different DN values present in the region of interest (ROI). As the maximum entropy value of a ROI could only be reached when the occurrence of the DNs values (*i.e.*, states) is equiprobable, the maximum entropy (H_{max}) is calculated considering all DNs values with the same probability as follows (Eq.(1)):

$$H_{max} = \log_2 N. \tag{1}$$

Dividing the number of pixels that have a determined DN value by the total DN values present in the ROI, we have the probability p of the i^{th} DN value of occurrence of this value within the ROI. The Boltzmann-Gibbs-Shannon entropy (H_e) for ROI is then calculated as (Eq.(2)):

$$H_e = -\sum_{DN \in N} P_{(DN)} \log_2 P_{(DN)}. \tag{2}$$

Finally, the variability measure (V) is obtained by dividing the information entropy calculated (H_e) by the maximum entropy (H_{max}), as follows (Eq.(3)):

$$V = \frac{H_e}{H_{max}}. \tag{3}$$

It can be deduced that complexity values for this measure range between 0 and 1, with complexity values associated with disorder (thermodynamic equilibrium).

Differently from the variability measure H_e/H_{max}, SDL and LMC belong to the third category of complexity measures defined by Shiner et al. [22], considering that the highest complexity is situated between order/homogeneous and disorder/heterogeneous patterns, that is, regions of intermediary heterogeneity associated with a high degree of self-organization. A convex function of information entropy may mathematically represent this assumption.

SDL measure is composed of two terms: disorder and order, i.e. (Eq.(4)):

$$SDL = (H_e/H_{max})[1 - (H_e/H_{max})]. \tag{4}$$

For the LMC measure, the order term is substituted by another term, called disequilibrium (D), which measures the distance between the system probability and the uniform distribution [27] (Eq.(5)):

$$D = \sum_{i=1}^{N} \left(P_{(DN)} - \frac{1}{N} \right)^2. \tag{5}$$

Consequently, LMC is given by (Eq.(6))

$$LMC = (H_e/H_{max})[1 - (H_e/H_{max})] \tag{6}$$

or by (Eq.(7)):

$$LMC = (H_e/H_{max})D \tag{7}$$

Zero is the minimum value for both measures, while 0.25 and 0.15 are the maximum values for SDL and LMC, respectively. These maximun values occur when the DN distribution is uniform [27].

To apply these measures based on information entropy to the remote sensing images, we developed two scripts in Phyton language to be executed as plugins in QGIS, an open-source Geographic Information System. The first one is CompPlex H_eROI, which calculates He/Hmax, LMC, and SDL complexity measures of a ROI and compares them with others patches and their transition areas. The other plugin is CompPlex Janus, composed of a sliding window that runs through the image, calculating those three measures for the set of pixels inside it. CompPlex Janus then generates complexity maps, allowing verification as complexity varies in the landscape over space and time.

Here we present examples illustrating the application of CompPlex H_eROI to evaluate the complexity of several ROIs (Example 1) and the use of CompPlex Janus to evaluate the spatial distribution of landscape patterns complexity (Example 2), highlighting in both cases the efficiency of the measures based on the information entropy presented.

2.1.1.1 Example 1: CompPlex H_eROI applied to evaluate patterns of different land uses

In this example, we show how CompPlex H_eROI had been applied to evaluate, by using metrics based on information entropy, the complexity of spatial patterns with different land uses present in two river neighbor basins located at municipality of São Carlos (São Paulo state, Brazil – **Figure 2**), especially as indicators of the resilience of its green areas, to help establishing a free space system for this region. Land uses inside these river basins had been identified using images from CBERS 4 remote sensor (**Figure 3**), and six categories of use were selected to be evaluated by CompPlex H_eROI. Results obtained for ROIs of these categories are shown in **Table 1**, where they are compared for each measure and each band used.

Colors associated with values presented in **Table 1** help to identify tendencies of each land use complexity pattern. In general, ROIs of exposed soil, urban areas, and pasture have high values for H_e, $H_{max,}$ and H_e/H_{max} measures and low values for SDL and LMC measures, indicating that these land uses have more disordered patterns. In turn, agricultural use varies from low to relatively high values for H_e, $H_{max,}$ and H_e/H_{max} measures, but has, in most cases, low values for SDL and LMC measures. Finally, vegetation areas have low values for the first three measures (H_e, H_{max} and H_e/H_{max}) and high values for SDL and LMC measures that use convex

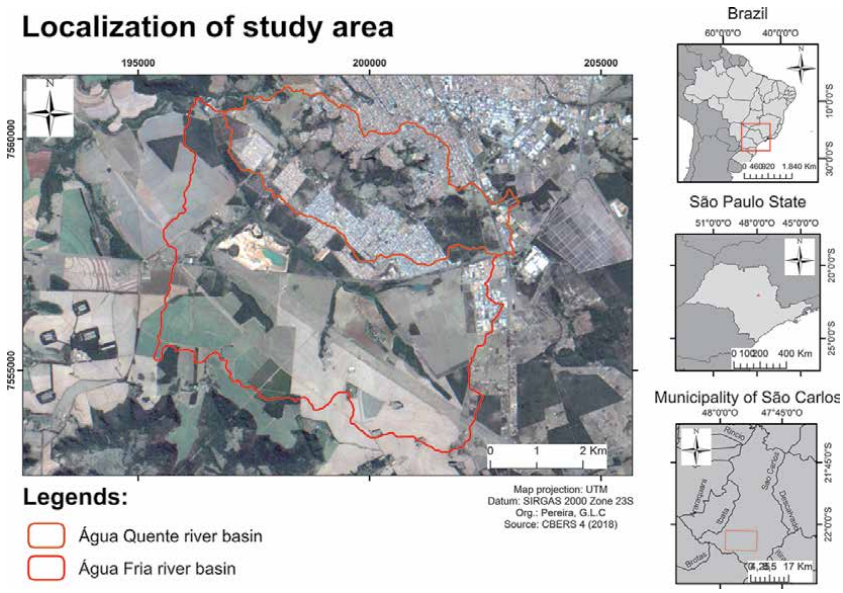


Figure 2.
Localization of study area in São Carlos's municipality (São Paulo state, Brazil).

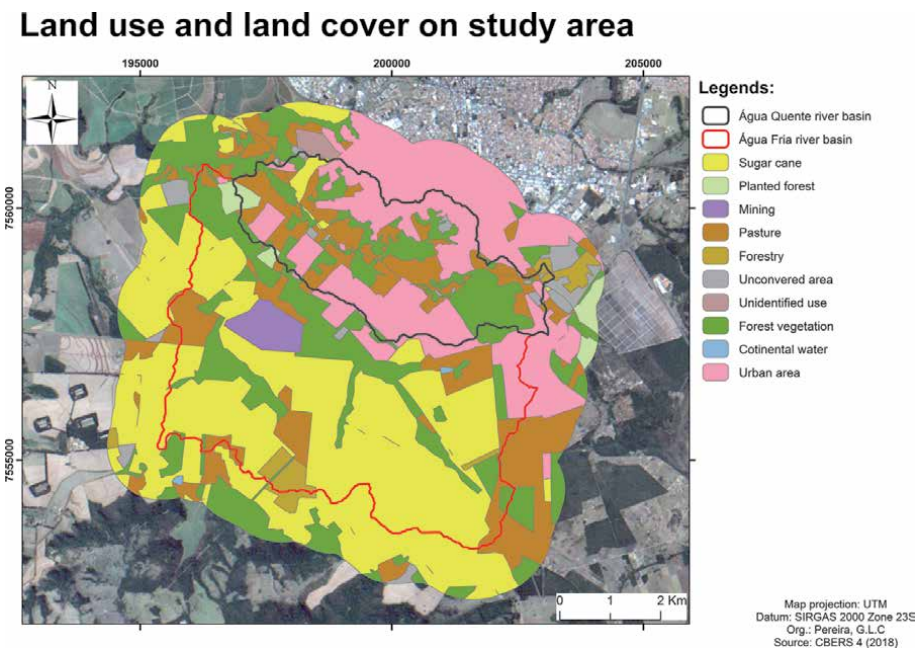


Figure 3.
Land use and land cover on study area (municipality of São Carlos, São Paulo state, Brazil).

function of disorder to associate more complexity with patterns situated in a zone between ordered and disordered patterns. Therefore, these results are coherent and consistent with the Landscape Ecology assumption that more complexity is found in intermediary heterogeneity patterns [5, 6], as is the case of vegetation areas present in the two neighbor river basins studied. High values for SDL and LMC obtained by these areas could be related to their high levels of self-organization and resilience.

Land use	ROI	He			Hmax			He/Hmax			SDL			LMC		
		Band 5	Band 6	Band 7	Band 5	Band 6	Band 7	Band 5	Band 6	Band 7	Band 5	Band 6	Band 7	Band 5	Band 6	Band 7
Exposed soil	ROI 1	4,4811	3,9741	3,5066	5,0444	4,6439	4,2479	0,8883	0,8558	0,8255	0,0992	0,1234	0,1441	0,0256	0,0366	0,0485
	ROI 2	4,1565	3,5415	3,0735	5,2095	4,5850	4,2479	0,7979	0,7724	0,7235	0,1613	0,1758	0,2000	0,0361	0,0527	0,0696
	ROI 3	4,2545	3,6888	3,2378	5,0444	4,3923	4,0000	0,8434	0,8398	0,8094	0,1321	0,1345	0,1542	0,0358	0,0504	0,0666
	ROI 4	4,0900	3,4612	2,9686	4,3923	3,9069	3,4594	0,9323	0,8859	0,8581	0,0631	0,1011	0,1218	0,0211	0,0413	0,0466
	ROI 5	3,3000	2,6311	2,5105	4,7549	4,1699	3,7004	0,6928	0,6310	0,6784	0,2128	0,2328	0,2182	0,0719	0,1054	0,0980
Urban area	ROI 1	4,6000	4,1853	3,8554	5,0875	4,7549	4,3923	0,9061	0,8802	0,8778	0,0851	0,1054	0,1073	0,0159	0,0228	0,0274
	ROI 2	5,0000	4,4723	4,0663	5,6724	5,3576	5,1293	0,8802	0,8348	0,7928	0,1054	0,1379	0,1643	0,0145	0,0234	0,0322
	ROI 3	4,7000	4,4006	4,1164	5,7004	5,7814	5,4263	0,8279	0,7612	0,7586	0,1425	0,1818	0,1831	0,0234	0,0315	0,0374
	ROI 4	5,1000	4,8281	4,5210	6,1293	6,1699	6,0444	0,8344	0,7825	0,7480	0,1382	0,1702	0,1885	0,0173	0,0253	0,0334
	ROI 5	6,0000	5,9472	5,6864	6,4919	6,5236	6,3399	0,9306	0,9116	0,8969	0,0646	0,0805	0,0924	0,0072	0,0099	0,0134
	ROI 6	4,3000	3,8912	3,4527	4,7004	4,5236	4,1699	0,9061	0,8602	0,8280	0,0851	0,1203	0,1424	0,0182	0,0319	0,0453
	ROI 7	4,8000	4,3885	4,0597	6,5546	6,4757	6,3750	0,7389	0,6777	0,6368	0,1929	0,2184	0,2313	0,0240	0,0355	0,0437
Forest vegetation	ROI 1	3,3000	2,8502	2,5804	4,4594	4,0000	3,7004	0,7327	0,7125	0,6973	0,1958	0,2048	0,2111	0,0869	0,1109	0,1092
	ROI 2	3,7000	3,2672	2,7327	4,6439	4,3219	3,8074	0,7944	0,7560	0,7177	0,1633	0,1845	0,2026	0,0570	0,0694	0,0904
	ROI 3	3,4000	2,8545	2,5770	4,9542	4,9069	4,5850	0,6779	0,5817	0,5621	0,2184	0,2433	0,2461	0,0684	0,0906	0,1017
	ROI 4	3,3000	2,8845	2,4381	4,7549	4,2479	3,8074	0,6880	0,6790	0,6404	0,2146	0,2179	0,2303	0,0815	0,0920	0,1112
	ROI 5	3,7000	3,3071	2,8197	4,8074	4,3923	4,1699	0,7684	0,7529	0,6762	0,1779	0,1860	0,2190	0,0532	0,0633	0,0895
	ROI 6	3,5000	3,2881	2,8041	4,5850	4,3923	3,9069	0,7631	0,7486	0,7177	0,1808	0,1882	0,2026	0,0571	0,0616	0,0835
	ROI 7	3,2000	2,8528	2,4169	4,8580	4,5236	4,2479	0,6535	0,6307	0,5690	0,2264	0,2329	0,2452	0,0793	0,0943	0,1117
	ROI 8	3,5000	3,0592	2,6443	4,9542	4,6439	4,0875	0,7122	0,6588	0,6469	0,2050	0,2248	0,2284	0,0659	0,0794	0,0950
	ROI 9	2,8000	2,5271	2,2851	4,3923	4,0875	3,5850	0,6276	0,6183	0,6374	0,2337	0,2360	0,2311	0,1100	0,1139	0,1247
	ROI 10	2,6000	2,2068	1,9570	3,8074	3,3219	2,8074	0,6828	0,6643	0,6971	0,2166	0,2230	0,2112	0,1168	0,1316	0,1251
	ROI 11	3,7000	3,1969	2,8469	4,9069	4,5850	4,3219	0,7505	0,6973	0,6587	0,1873	0,2111	0,2248	0,0760	0,0979	0,1146
	ROI 12	2,8000	2,3646	2,2609	4,1699	3,5850	3,4594	0,6766	0,6596	0,6535	0,2188	0,2245	0,2264	0,1153	0,1334	0,1272
	ROI 13	3,0000	2,6754	2,3351	4,1699	3,9069	3,7004	0,7188	0,6848	0,6310	0,2021	0,2159	0,2328	0,0757	0,1000	0,1121
	ROI 14	2,9000	2,6716	2,2181	4,2479	3,9069	3,4594	0,6725	0,6838	0,6412	0,2202	0,2162	0,2301	0,0963	0,1000	0,1232
	ROI 15	2,4000	2,0496	1,9656	4,0875	3,5850	3,0000	0,5805	0,5717	0,6552	0,2435	0,2449	0,2259	0,1433	0,1516	0,1380
	ROI 16	2,7000	2,6331	2,0771	3,9069	3,7004	3,3219	0,6934	0,7116	0,6253	0,2126	0,2052	0,2343	0,0902	0,0926	0,1244
	ROI 17	3,6190	3,2078	2,9394	4,9542	4,5236	4,0875	0,7305	0,7091	0,7191	0,1969	0,2063	0,2020	0,0837	0,0955	0,0985
	ROI 18	2,6816	2,1261	2,0039	4,2479	3,7004	3,4594	0,6313	0,5745	0,5793	0,2328	0,2444	0,2437	0,1124	0,1364	0,1429
	ROI 19	2,2629	1,8992	1,8473	3,4594	3,1699	2,8074	0,6541	0,5991	0,6580	0,2262	0,2402	0,2250	0,1148	0,1474	0,1311
	ROI 20	3,5891	3,1214	2,7833	4,7004	4,0875	3,4594	0,7636	0,7637	0,8045	0,1805	0,1805	0,1573	0,0778	0,0890	0,0853
	ROI 21	3,6965	3,2385	2,6494	4,2479	3,7004	3,1699	0,8702	0,8752	0,8358	0,1130	0,1093	0,1372	0,0403	0,0448	0,0693
	ROI 22	2,5271	1,9190	1,8315	3,8074	3,1699	3,1699	0,6637	0,6054	0,5778	0,2232	0,2389	0,2440	0,1287	0,1422	0,1488
	ROI 23	3,1657	2,8982	2,2734	4,1699	3,7004	3,1699	0,7592	0,7832	0,7172	0,1828	0,1698	0,2028	0,0923	0,0805	0,1129
Agriculture	ROI 1	3,4584	2,9410	2,3150	4,0875	3,7004	3,1699	0,8461	0,7948	0,7303	0,1302	0,1631	0,1970	0,0453	0,0621	0,1011
	ROI 2	3,6998	3,3780	2,6080	4,0875	3,8074	3,1699	0,9052	0,8872	0,8227	0,0858	0,1001	0,1458	0,0294	0,0366	0,0699
	ROI 3	4,0709	3,6021	2,9829	5,0000	4,7004	4,2479	0,8142	0,7663	0,7022	0,1513	0,1791	0,2091	0,0375	0,0451	0,0687
	ROI 4	4,7789	4,0867	3,7128	5,6724	5,5546	5,4594	0,8425	0,7357	0,6801	0,1327	0,1944	0,2176	0,0171	0,0338	0,0416
	ROI 5	3,2443	2,6515	2,1540	4,3923	3,9069	3,3219	0,7386	0,6787	0,6484	0,1931	0,2181	0,2280	0,0621	0,0995	0,1225
	ROI 6	4,3082	3,7854	3,2445	5,3219	5,0000	4,5236	0,8095	0,7571	0,7172	0,1542	0,1839	0,2028	0,0298	0,0427	0,0597
Mining	ROI 1	5,8503	5,2216	4,5579	6,3038	6,0000	5,4919	0,9281	0,8703	0,8299	0,0668	0,1129	0,1411	0,0070	0,0142	0,0250
Pasture	ROI 1	4,3831	3,6957	3,2239	5,3576	5,0000	4,6439	0,8181	0,7391	0,6942	0,1488	0,1928	0,2123	0,0253	0,0442	0,0577
	ROI 2	3,9380	3,5371	2,9939	4,6439	4,2479	3,9069	0,8480	0,8327	0,7663	0,1289	0,1393	0,1791	0,0297	0,0396	0,0595
	ROI 3	4,0871	3,6439	3,0653	4,6439	4,0875	3,5850	0,8801	0,8915	0,8550	0,1055	0,0967	0,1239	0,0237	0,0253	0,0408
	ROI 4	4,2984	3,6619	3,1821	5,0000	4,5850	4,4594	0,8597	0,7987	0,7136	0,1206	0,1608	0,2044	0,0249	0,0418	0,0615
	ROI 5	4,5809	4,1065	3,6705	5,6724	5,4919	5,1293	0,8076	0,7477	0,7156	0,1554	0,1886	0,2035	0,0269	0,0397	0,0523
	ROI 6	4,0211	3,4431	3,0533	5,0000	4,3923	3,8074	0,8042	0,7839	0,8019	0,1575	0,1694	0,1588	0,0389	0,0532	0,0585
	ROI 7	4,3997	4,0811	3,5683	4,6439	4,3219	3,9069	0,9474	0,9443	0,9133	0,0498	0,0526	0,0791	0,0127	0,0139	0,0232
	ROI 8	4,4676	3,9592	3,4447	5,3576	5,0000	4,7004	0,8339	0,7918	0,7328	0,1385	0,1648	0,1958	0,0219	0,0337	0,0497
	ROI 9	4,1668	3,5807	3,0151	4,5850	4,0000	3,5850	0,9088	0,8952	0,8410	0,0829	0,0938	0,1337	0,0219	0,0329	0,0533
	ROI 10	4,2233	3,7916	3,1975	4,5236	4,0875	3,5850	0,9336	0,9276	0,8919	0,0620	0,0671	0,0964	0,0135	0,0206	0,0323
	ROI 11	3,8417	3,4896	2,9152	4,3219	3,9069	3,3219	0,8889	0,8932	0,8775	0,0988	0,0954	0,1075	0,0284	0,0332	0,0463

Legends:
- for H_e, H_{max} and H_e/H_{max} : = high value = low value
- for SDL and LMC: = low value = high value

Table 1.
Results obtained by CompPlex HeROI for six land use and land cover categories in the study area (municipality of São Carlos, São Paulo state, Brazil).

2.1.1.2 Example 2: CompPlex Janus and landscape complexity maps

To exemplify how CompPlex Janus works to generate landscape complexity maps, here is presented a case study of sensor images from the Assis Ecological Station and its boundaries (located at Assis, São Paulo State, Brazil – **Figure 4**). Several tests varying the sliding window size, sensor band, and number of color classes had been performed to compare results obtained by H_e, H_e/H_{max}, SDL, and LMC measures. Some of these results are shown in **Figures 5** and **6**.

Comparing the four examples of maps of H_e, we observe that for sliding window of 3x3 pixels sides (**Figure 5A** and **B**), this measure highlights borders among different land uses and, especially for the image of band 3 (**Figure 5A**), the apparent homogeneity of natural vegetation is broken. For this same measure, but for sliding window of 9x9 pixels sides results are shown in **Figure 5C** and **D**. On the other side, these edges are blurred for the 9x9 pixel window. However, areas with higher values for this measurement (visualized by more intense red tones) are found around natural vegetation areas, possibly indicating areas of greater risk to their integrity and resilience.

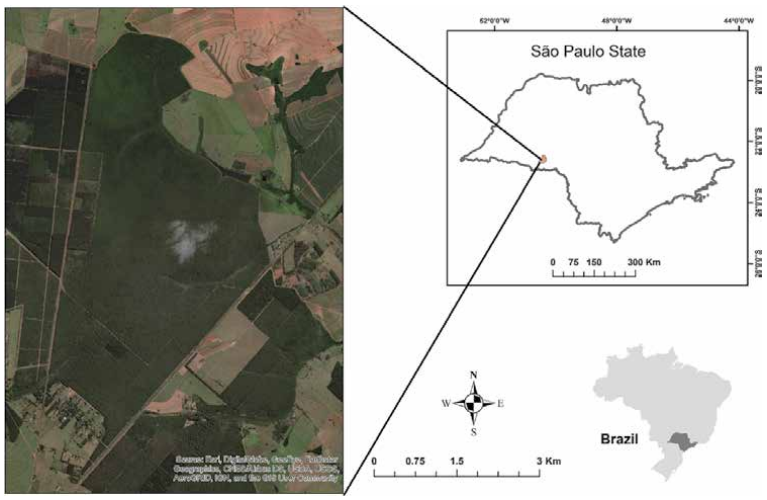


Figure 4.
Localization of Assis Ecological Station and surroundings (municipality of Assis, State of São Paulo, Brazil).

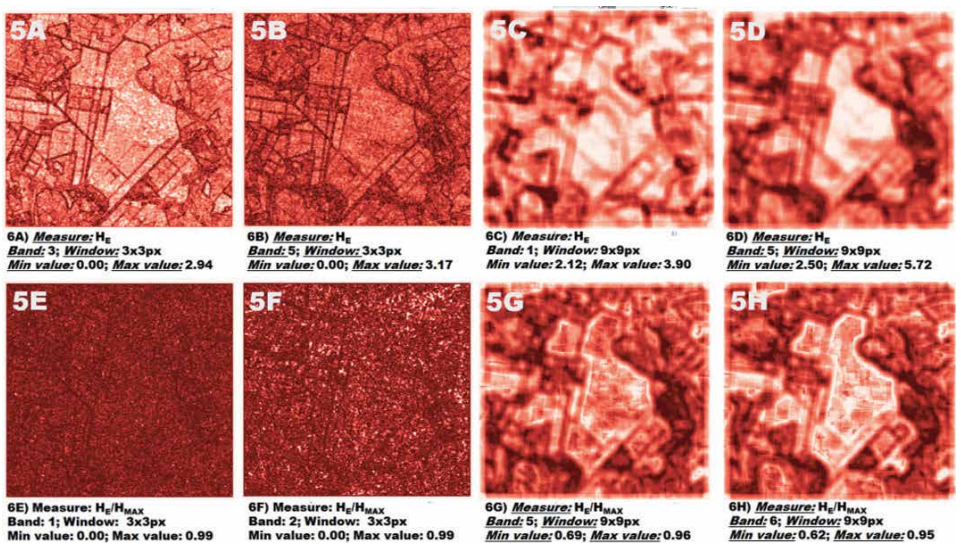


Figure 5.
Some complexity maps for He and H_e/H_{max} entropy measures generated by CompPlex Janus to study area.

For H_e/H_{max} measure, we can perceive a significant difference between maps generated by a window of 3x3 pixels sides (**Figure 5E** and **F**) and those of 9x9 pixels (**Figure 5G** and **H**). Due to the reduced amount of pixels in the smaller window, there is less diversity of information, and the interval between minimum and maximum values is high. As occurred with the He measurement, a larger window (9x9 pixels) generated a more extensive range for the minimum and maximum values, highlighting possible areas that represent greater risks to natural vegetation.

In **Figure 6A-H**, we show, respectively, some results obtained by Complex Janus to SDL and LMC measure. For a window of 3x3 pixels sides (**Figure 6A, B, E** and **F**), these measures allow identifying punctual areas with higher values within natural vegetation regions. This effect is best observed on images generated by windows of

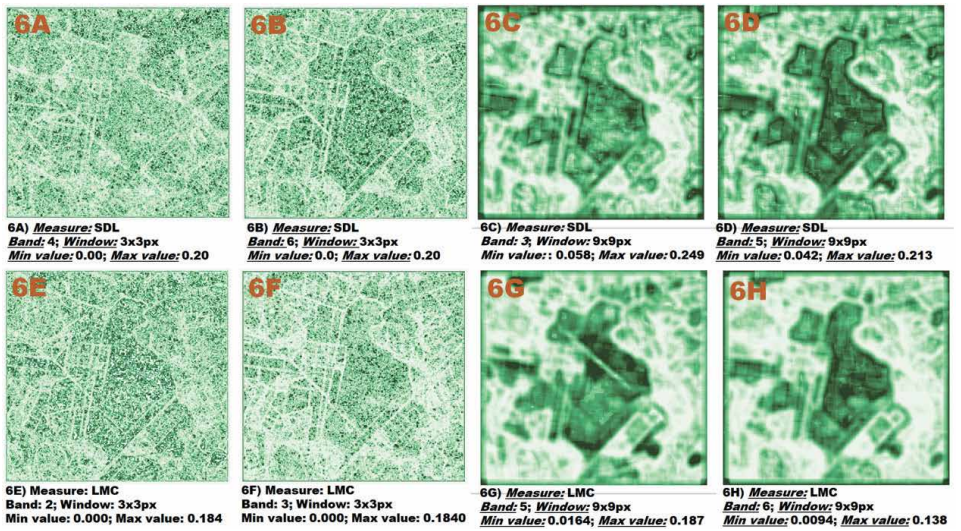


Figure 6.
Some complexity maps for SDL and LMC entropy measures generated by CompPlex Janus to study area.

9x9 pixels (**Figure 6C, D, G**, and **H**), where areas of natural vegetation have a more variable value gradient and areas with other land uses are 'homogenized' with low values for both measures. In general, SDL and LMC measures assign higher values to natural vegetation, consistent with the assumption of Landscape Ecology that more complex patterns are associated with intermediary spatial heterogeneity.

3. Measures based on information entropy applied to analyze climatic time series

Information entropy measures are also useful to verify complexity (in the sense of variability) of time series, as shown by Piqueira and Mattos [9]. To exemplify how these measures can be utilized for this purpose, here we show an application of H_e/H_{max}, SDL, and LMC measures for a time series corresponding to the maximum daily temperatures that occurred in each January from 1980 to 2017 in the municipality of São Carlos (São Paulo State, Brazil).

To calculate the measures, daily maximum temperature data for each January of the entire time series were used to define its quartiles. Then, for the January data for each year, we check the number of days that belonged to each quartile, which allowed us to calculate the probability p for each interval. The system extension (N) corresponded to the number of quartiles that presented at least one data.

With these values, it was possible to calculate the measures H_e/H_{max}, SDL, and LMC for data from January of each year, according to the equations previously presented. The results obtained are shown in **Table 2**. To compare the performance of measures and to try to identify any patterns from results for each measure we group each of them in decreasing order of value and organize in four classes. **Table 3** shows pairwise comparisons between the measures to verify whether or not a given year occupies the same class and the same position for both measures. Through these comparisons, it is evident that H_e/H_{max} and SDL have the same behavior, while LMC measure behavior differs from them, revealing the differences in the relations between order and disorder terms present in these measures' equations.

Month/Year	He/Hmax	SDL	LMC	El Niño
jan/80	0,996121	0,003864	0,002764	Moderate
jan/81	0,891832	0,096468	0,058234	-
jan/82	0,842241	0,132871	0,104075	-
jan/83	0,955452	0,042563	0,028501	Strong
jan/84	0,883083	0,103247	0,083392	-
jan/85	0,899002	0,090798	0,056753	-
jan/86	0,929433	0,065587	0,041346	-
jan/87	0,943894	0,052958	0,036096	Moderate
jan/88	0,913992	0,078611	0,057778	Strong
jan/89	0,807586	0,155391	0,098112	-
jan/90	0,898222	0,091419	0,053043	-
jan/91	0,888074	0,099399	0,072774	-
jan/92	0,946830	0,050343	0,038179	Strong
jan/93	0,987089	0,012744	0,008988	Weak
jan/94	0,887049	0,100193	0,059767	-
jan/95	0,958728	0,039569	0,028682	-
jan/96	0,975916	0,023504	0,017010	-
jan/97	0,750346	0,187327	0,128636	-
jan/98	0,815862	0,150231	0,111003	Strong
jan/99	0,959517	0,038844	0,024712	-
jan/00	0,982448	0,017244	0,010990	-
jan/01	0,770428	0,176869	0,128872	-
jan/02	0,995932	0,004051	0,002850	-
jan/03	0,958728	0,039569	0,028682	Moderate
jan/04	0,882989	0,103319	0,063169	-
jan/05	0,944311	0,052588	0,030216	-
jan/06	0,906762	0,084544	0,059208	-
jan/07	0,933882	0,061746	0,037657	Moderate
jan/08	0,920452	0,073220	0,048609	-
jan/09	0,869816	0,113236	0,087570	-
jan/10	0,933882	0,061746	0,037657	Moderate
jan/11	0,987089	0,012744	0,008988	-
jan/12	0,872595	0,111173	0,076878	-
jan/13	0,964841	0,033923	0,024765	-
jan/14	0,987089	0,012744	0,008988	-
jan/15	0,934765	0,060979	0,119966	-
jan/16	0,989680	0,010213	0,006951	Strong
jan/17	0,998848	0,001150	0,000780	-

For H_e/H_{max}: = low value = high value

For SDL and LMC: = low value = high value

Table 2.
Results obtained by applying entropy measures for daily maximum temperature data of municipality of São Carlos (São Paulo state, Brazil) for each January from 1980 to 2017.

Group	Comparaison		Comparaison		Comparaison	
	He/Hmax*	SDL**	He/Hmax*	LMC**	SDL**	LMC**
1	jan/97	jan/97	jan/97	jan/01	jan/97	jan/01
	jan/01	jan/01	jan/01	jan/97	jan/01	jan/97
	jan/89	jan/89	jan/89	jan/15	jan/89	jan/15
	jan/98	jan/98	jan/98	jan/98	jan/98	jan/98
	jan/82	jan/82	jan/82	jan/82	jan/82	jan/82
	jan/09	jan/09	jan/09	jan/89	jan/09	jan/89
	jan/12	jan/12	jan/12	jan/09	jan/12	jan/09
	jan/04	jan/04	jan/04	jan/84	jan/04	jan/84
	jan/84	jan/84	jan/84	jan/12	jan/84	jan/12
	jan/94	jan/94	jan/94	jan/91	jan/94	jan/91
2	jan/91	jan/91	jan/91	jan/04	jan/91	jan/04
	jan/81	jan/81	jan/81	jan/94	jan/81	jan/94
	jan/90	jan/90	jan/90	jan/06	jan/90	jan/06
	jan/85	jan/85	jan/85	jan/81	jan/85	jan/81
	jan/06	jan/06	jan/06	jan/88	jan/06	jan/88
	jan/88	jan/88	jan/88	jan/85	jan/88	jan/85
	jan/08	jan/08	jan/08	jan/90	jan/08	jan/90
	jan/86	jan/86	jan/86	jan/08	jan/86	jan/08
	jan/07	jan/07	jan/07	jan/86	jan/07	jan/86
	jan/10	jan/10	jan/10	jan/92	jan/10	jan/92
3	jan/15	jan/15	jan/15	jan/07	jan/15	jan/07
	jan/87	jan/87	jan/87	jan/10	jan/87	jan/10
	jan/05	jan/05	jan/05	jan/87	jan/05	jan/87
	jan/92	jan/92	jan/92	jan/05	jan/92	jan/05
	jan/83	jan/83	jan/83	jan/95	jan/83	jan/95
	jan/95	jan/95	jan/95	jan/03	jan/95	jan/03
	jan/03	jan/03	jan/03	jan/83	jan/03	jan/83
	jan/99	jan/99	jan/99	jan/13	jan/99	jan/13
	jan/13	jan/13	jan/13	jan/99	jan/13	jan/99
4	jan/96	jan/96	jan/96	jan/96	jan/96	jan/96
	jan/00	jan/00	jan/00	jan/00	jan/00	jan/00
	jan/93	jan/93	jan/93	jan/93	jan/93	jan/93
	jan/11	jan/11	jan/11	jan/11	jan/11	jan/11
	jan/14	jan/14	jan/14	jan/14	jan/14	jan/14
	jan/16	jan/16	jan/16	jan/16	jan/16	jan/16
	jan/02	jan/02	jan/02	jan/02	jan/02	jan/02
	jan/80	jan/80	jan/80	jan/80	jan/80	jan/80
	jan/17	jan/17	jan/17	jan/17	jan/17	jan/17

belongs to same group and occupy the same position

belongs to same group but occupy different position

doesn´t belong to the same group

* From minor value to major value

** From major value to minor value

Table 3.
Results obtained by applying entropy measures for daily maximum temperature data of municipality of São Carlos (São Paulo state, Brazil) for each January from 1980 to 2017.

4. Conclusions

In addition to exploring results obtained from case studies here presented, we intended to show why and how measures based on information entropy can contribute to understanding complexity of landscapes patterns and processes. As shown in the first example, H_e/H_{max}, SDL, and LMC are complexity measures that represent useful tools for evaluating landscape patterns. H_e/H_{max} allows identifying ordered and disordered targets, while SDL and LMC are related to intermediary heterogeneity patterns presented by landscape patches. Comparing the landscape metrics used here with the spectral decomposition methods proposed in Mustard and Sunshine [13], they prove to be quite efficient in comparing the complexity of the patterns of different patches as well as their variation over the entire landscape. Based on this example, the application of such metrics is proposed for multitemporal studies of landscape dynamics, for evaluating resilience and the degree of degradation of different fragments, for estimating the degree of the anthropic impact due to alterations on land usage, among other applications.

In the second example, we highlight the use of these measures to evaluate complexity in climatic time series. Our future studies involve the application of these measures as alternatives for classical statistical analysis, using them to assess the influences of both natural processes, such as El Niño and La Niña, and those resulting from anthropic processes, such as the increase in temperature and frequency of extreme weather events, such as severe droughts and heavier rains.

Author details

Sérgio Henrique Vannucchi Leme de Mattos[1*], Luiz Eduardo Vicente[2], Andrea Koga Vicente[2], Cláudio Bielenki Junior[1], Maristella Cruz de Moraes[1], Gabriele Luiza Cordeiro[1] and José Roberto Castilho Piqueira[3]

1 Environmental Complex Systems Laboratory, Department of Hydrobiology, Biological and Health Sciences Center, Federal University of São Carlos (UFSCar), São Carlos, SP, Brazil

2 ABC Platform, Embrapa Environment, Jaguariúna, SP, Brazil

3 Department of Telecommunications and Control Engineering, Polytechnic School of the University of Sao Paulo (USP), São Paulo, SP, Brazil

*Address all correspondence to: sergiomattos@ufscar.br

References

[1] MORIN, E. **From the concept of system to the paradigm of complexity**. Journal of Social and Evolutionary Systems. Volume 15, Issue 4, 1992, Pages 371-385.

[2] CHRISTOFOLETTI, A. **Modelagem de Sistemas Ambientais**. Editora Blucher, 1999. 256 Pages.

[3] KOESTLER, A. **The Ghost in the Machine**. Macmillan Length. 384 Pages.

[4] TURNER, M.G. & GARDNER, R.H. **Quantitative methods in landscape ecology: an introduction.** In: _. *Quantitative methods in landscape ecology*. New York: Springer Verlag, 1990. p.3-14.

[5] ANAND, M. et al. **Ecological Systems as Complex Systems: Challenges for an Emerging Science**. Diversity 2010, 2, 395-410; doi:10.3390/d2030395.

[6] PARROTT, L; **Measuring ecological complexity**. Ecological Indicators, November 2010, 10(6):1069-1076. doi: 10.1016/j.ecolind.2010.03.014.

[7] MATTOS, Sérgio Henrique Vannucchi Leme de; PIQUEIRA, José Roberto Castilho ; VASCONCELLOS NETO, João ; ORSATTI, Fernando Moya. **Measuring q-bits in three-trophic level systems**. Ecological Modelling, v. 200, n.1-2, p. 183-187, 2007.

[8] PIQUEIRA, J.R,C., MATTOS, S.H.V. L; VASCONCELOS NETO, J. **Measuring complexity in three-trophic level systems. Ecological Modelling**, v. 220, p. 266-271, 2009.

[9] PIQUEIRA, J. R. C.; MATTOS, S.H.V. L. **LMC and SDL complexity measures: a tool to explore time series**. Complexity, v. 2019, p. 1-8, 2019.

[10] AVERY, T.E., BERLIN, G.L. **Fundamentals of remote sensing and airphoto interpretation.** New York: Macmillan, 1992. 472p.

[11] JONES, H.G., VAUGHAN, R.A. **Remote Sensing of Vegetation: Principles, Techniques, and Applications**. UP Oxford, 2010. 384 Pages.

[12] MUSICK, H.B. & GROVER, H.D. **Image textural measures as indices of landscape pattern**. In: Turner, M.G. & Gardner, R.H. Quantitative methods in landscape ecology. New York: Springer Verlag, 1990. p.77-103.

[13] MUSTARD, J.F.; SUNSHINE, J.M. **Spectral analysis for Earth Science: investigations using remote sensing data.** In: RENCZ, A.N. Remote sensing for the Earth Sciences. New York: John Wiley & Sons, 1999.

[14] ASNER, G.P. et al. **Remote sensing of selective logging in Amazonia-Assessing limitations based on detailed field observations, Landsat ETM+, and textural analysis**. Remote Sensing of Environment, v.80, 2002, p. 483-496.

[15] GARRIGUES, S. *et al.* **Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data.** Remote Sensing of Environment, v.105, n.4.2006, p.286-298.

[16] MIRANDA, F.P. *et al.* **Analysis of JERS-1 (Fuyo-1) SAR data for vegetation discrimination in northwestern Brazil using the semivariogram textural classifier (STC).** *International Journal of Remote Sensing*, v.17, 1996.

[17] BAK, P. **How nature works: the science of self-organized criticality.** New York: Springer-Verlag, 1997. 212p.

[18] FARINA, A. **Principles and methods in landscape ecology.** Londres: Chapman & Hall, 1998. 235p.

[19] LI, B.L. **Fractal geometry applications in description and analysis of patch patterns and patch dynamics.** Ecological Modelling, v.132, 2000. p.33-50.

[20] NAVEH, Z. & LIEBERMAN, A. **Landscape ecology.** New York: Springer Verlag, 1994. 360p.

[21] FEAGIN, R.A. **Heterogeneity versus homogeneity: a conceptual and mathematical theory in terms of scale-invariant and scale-covariant distributions.** Ecological Complexity, v.2, 2005, p.339–356.

[22] SHINER, J.S. et al. **Simple measure of complexity.** Physical Review, 1999, v.59, n.2, Pages 1459-1464.

[23] KANEKO, K., TSUDA, I. **Complex Systems: chaos and beyond.** Berlim: Springer Verlag, 2001. 257p.

[24] PARROT, L. **Quantifying the complexity of simulated spatiotemporal population dynamics.** Ecological Complexity, v.2, 2005, p.175-184.

[25] LÓPEZ-RUIZ, R. *et al.* A statistical measure of complexity. **Physical Letter A**, v.209, 1995, p.321-326.

[26] SHANNON, C. E. **A mathematical theory of communication**. The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

[27] PIQUEIRA, J.R,C. **A comparison of LMC and SDL complexity measures on binomial distributions.** Physica. A (Print), v. 444, p. 271-275, 2016.

Chapter 5

Optimization of the ANNs Models Performance in the Short-Term Forecasting of the Wind Power of Wind Farms

Sergio Velázquez-Medina and Ulises Portero-Ajenjo

Abstract

Due to the low dispatchability of wind power, the massive integration of this energy source in electrical systems requires short-term and very short-term wind farm power output forecasting models to be as efficient and stable as possible. A study is conducted in the present paper of potential improvements to the performance of artificial neural network (ANN) models in terms of efficiency and stability. Generally, current ANN models have been developed by considering exclusively the meteorological information of the wind farm reference station, in addition to selecting a fixed number of time periods prior to the forecasting. In this respect, new ANN models are proposed in this paper, which are developed by: varying the number of prior 1-h periods (periods prior to the prediction hour) chosen for the input layer parameters; and/or incorporating in the input layers data from a second weather station in addition to the wind farm reference station. It has been found that the model performance is always improved when data from a second weather station are incorporated. The mean absolute relative error (MARE) of the new models is reduced by up to 7.5%. Furthermore, the longer the forecast horizon, the greater the degree of improvement.

Keywords: Artificial neural networks (ANN), wind power forecasting, model performance, wind farm power output

1. Introduction

A major impediment to the large-scale integration of wind power in electrical systems is the low dispatchability of this energy source. The effects of variations in wind speed, and hence wind power, are not only observed on a year-to-year or season-to-season scale, but also on a within-day scale [1–5]. A strategy that can be employed to improve wind energy integration in electrical systems is to optimize the performance of short-term forecasting models of wind farm power production. This strategy is the focus of the present study.

The direct consequences of the low dispatchability of wind power on electrical systems can be both technical and economic. Supply and demand adjustments in electrical systems are made 24–36 hours in advance. Any mismatches that might arise between supply and demand forecasting are subsequently corrected on the day

itself [6–9]. The mismatch correction as the result of imprecise forecasting entails additional costs for the electrical system [7, 10]. These extra costs are generally absorbed by the end user and/or electricity producer, with the latter thus burdened by an additional production cost.

Other strategies have been used to minimize the problem described above. One involves the direct estimation of the net energy demand of the electrical system, which can be understood as the difference between total demand and the energy generated by renewable sources. In [11–12], a model is proposed for direct forecasting of net energy demand which is validated with data from different electrical systems. Reference [13] compares a direct forecasting model of net energy demand with different indirect forecasting strategies.

In the electricity market, the matching of supply and demand is generally performed for 1 h periods. For this reason, in an analysis of model forecasting performance, it is very important to evaluate the error for 1 h periods, to study model performance for different forecast horizons, and to evaluate the stability of the error in the time horizon in which the forecasting is made.

Numerous studies can be found in the literature on the development of short-term forecasting models. Different techniques and approaches have been analyzed and proposed. In most cases, good performances for specific forecasting horizons have been obtained. The techniques that have been used range from simple heuristics [14–20] to systems which employ artificial intelligence [21–34]. The study developed in the present paper focuses on models which employ the technique of artificial neural networks (ANNs) to forecast wind farm power production [21, 22, 26], [27, 29–31, 33, 34].

In [34], the proposed forecasting model is developed on the basis of improvements made to the kriging interpolation method and empirical mode decomposition, using a new forecasting engine based on neural networks. To analyze the results, the mean absolute percentage error (MAPE), normalized mean absolute error (NMAE) and normalized root mean square error (NRMSE) metrics are used, calculated as the mean value in the forecasting horizons (24 h and 6 h). As in [34], models have been developed for different forecasting horizons [26, 27, 33]. However, an extensive analysis of the literature conducted by the authors of the present study has found that the models developed to date only consider a specific and fixed number of prior 1-*h* periods (periods prior to the prediction hour). It should also be noted that, in all the studies consulted, the meteorological data used as input layer parameters correspond exclusively to the reference weather station (WS) of the wind farm. In no case is the meteorological information used from additional WSs other than the reference WS of the wind farm. Finally, the metrics used to assess model performance in all these studies are obtained as the mean value of the forecasting time horizon. As previously stated, given that the matching of supply and demand in the electricity market is performed for 1 h periods, there is an additional interest in the study of the possible variation of the metrics within that time frame for each of the hourly periods.

The present study considers possible improvements, in terms of efficiency and stability, to the performance of ANN-based models for wind power forecasting. For this purpose, an analysis is made on the improvement of model performance of: ① varying the number of prior 1-*h* periods (periods prior to the forecasting hour) chosen for the ANN input layer parameters; and/or ② incorporating in the input layer data from a second weather station in addition to the data from the wind farm reference station. The analysis is undertaken for a wide range of forecasting horizons. Based on the above, a total of up to 175 ANN models are generated, and the results are compared by applying the models to two actual wind farms located in the Canary Islands, Spain.

The aim of this paper is to make the following original contributions to the scientific body of knowledge:

1. A study of improvement in the efficiency and stability of ANN models of varying the number of 1-*h* prior periods (periods prior to the prediction hour and hereinafter referred to as *n*), chosen for incorporation of the input layer parameters.

2. A study of improvement in ANN model performance of the additional incorporation in the input layer of meteorological data from WSs other than the wind farm reference station.

Both effects are analyzed for different forecasting horizons.

2. Methodology

Figure 1 shows the methodology followed in the present study for the implementation of different ANN models generated. It shows the combination of parameters which are considered for the input and output layer neurons in the generation process of different ANN models. The various parameters are defined as follows: t_i is the time instant on the basis of which the forecast is made, and V_{ti}, D_{ti} and P_{ti} are the wind speed, wind direction and the wind farm power output, respectively, in the instant t_i.

The following data are used in all the models: historical wind speed and direction data obtained from the wind farm reference WS, and historical power production

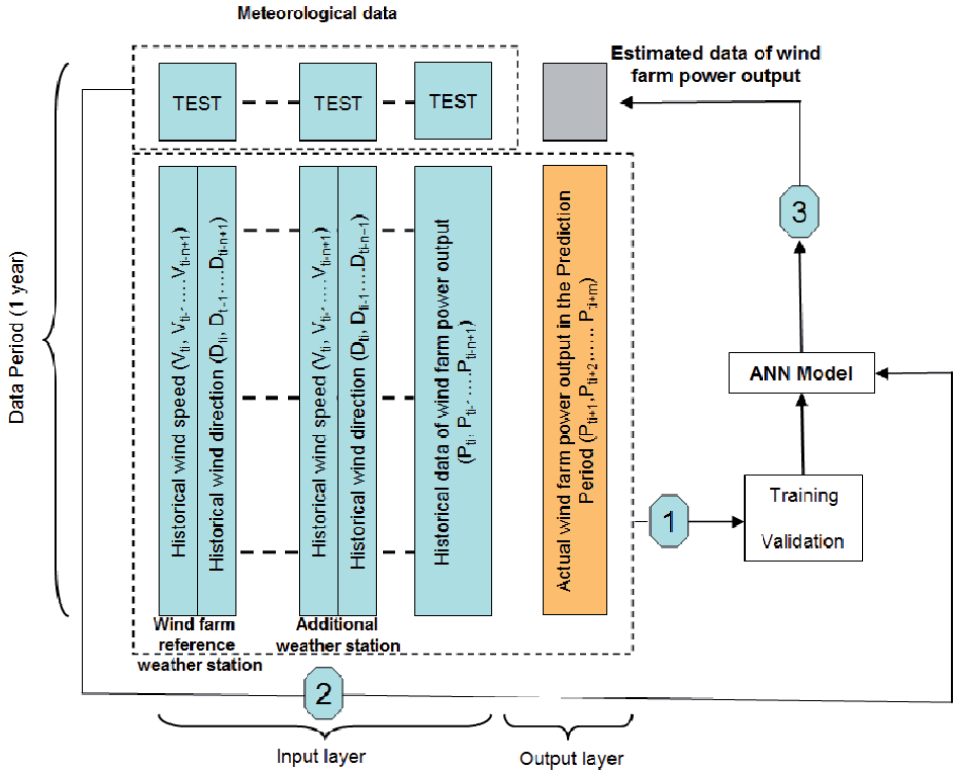


Figure 1.
Methodology to obtain forecasting models.

data of the wind farm. In some models, as will subsequently be explained, the historical wind speed and direction data of a second WS are used in addition to the data of the wind farm reference station.

The output layer is comprised of the power output values for different forecasting horizons.

The number of hours prior to the prediction hour, n, and the length of the forecasting horizon that is being forecasted, m, are variable.

2.1 Architecture of ANN employed

The ANNs used to generate the models are comprised of three layers with feedforward connections. For this purpose, multi-layer perceptron (MLP) topologies have been used [35, 36]. In order not to increase the length of the training period excessively, a single layer of hidden neurons is used. This architecture has been shown to have the capacity to satisfactorily approximate any continuous transformation [35, 36]. Various prior tests have been carried out to choose the number of hidden neurons, varying the number of input signals. It is found that using more than 20 neurons merely increases the time required for model training and validation without improving the results. It is therefore decided to use a total of 20 neurons in the hidden layer.

The architectures are trained using the backpropagation algorithm with sigmoidal activation function [31, 32]. The Levenberg–Marquardt algorithm is used to minimize the mean square error committed in the learning process [35, 37].

To carry out the training and validation stages used to generate the model and the test stage of the network, the available annual data series for each parameter are divided into random and different subsets (**Figure 1**). The proportion of data selected for each of the stages is 75%, 15% and 10%, respectively.

As can be seen in **Figure 1**, the training and validation data subsets are used to generate the model. The test data subset is used to evaluate the performance of the model generated.

The 10-fold cross-validation technique is used for the process of model generation and evaluation. The test stage data subset is used in each of the iterations. The error assigned to each model is the arithmetic mean of those obtained in the test stage for each of the iterations.

The various studies are performed using neural network tools available in the MATLAB software package.

2.2 Study cases

1. *Case A: Comparison of efficiency and stability of different ANN models obtained when varying the number of periods prior to the prediction hour (n) chosen for incorporation of different parameters in the input layer*

The number of prior periods, n, and the number of forecast horizon periods, m, are study variables. The different combinations of n and m generate different models whose performances will be analyzed. For Case A, both n and m are permitted to take the values 3, 6, 12, 24 and 36. That is to say, five different models are generated for each forecasting horizon, and thus the total number of generated models is 25. This methodology is applied to the two wind farms of the study.

To study the models in terms of the stability of forecasting, the results obtained for each of the periods within the forecasting horizon, m, are compared.

Figure 2 shows the structure of the neural network for this study case. The number of neurons of the output layer depends on the forecasting horizon, and will thus

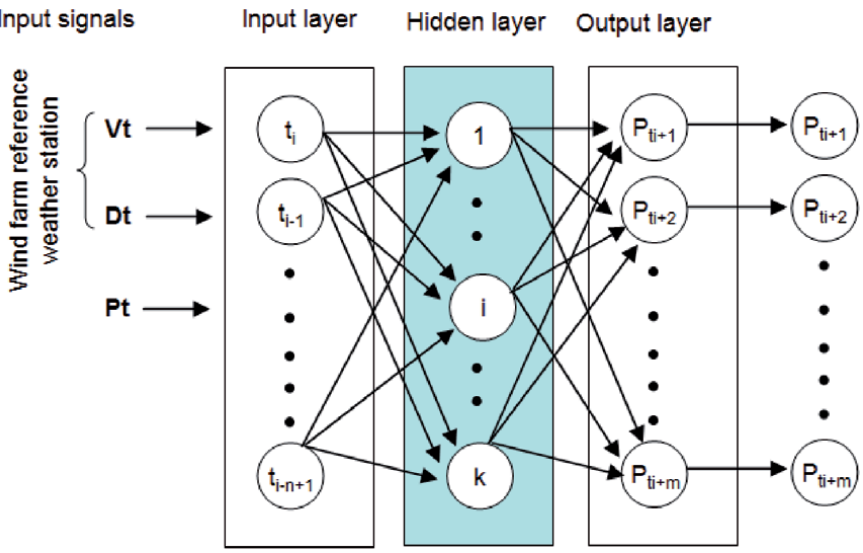


Figure 2.
Schematic representation of neural network for generation of forecasting models in case A.

fluctuate between 3 and 36 neurons. For the input layer, the number of neurons will also vary depending on the value of n, from 9 (n = 3) to 108 (n = 36) neurons.

2. *Case B: Comparison of performance of ANN models when additionally incorporating in the input layer the data from a second WS other than the reference station of wind farm.*

For Case B, both n and m could take the same values as indicated for Case A.

Figure 3 shows the structure of the neural network for the generation of models in Case B.

In Case B, the input layer of the ANN incorporates the data from a second WS in addition to that of the reference WS of the wind farm. To generate different models,

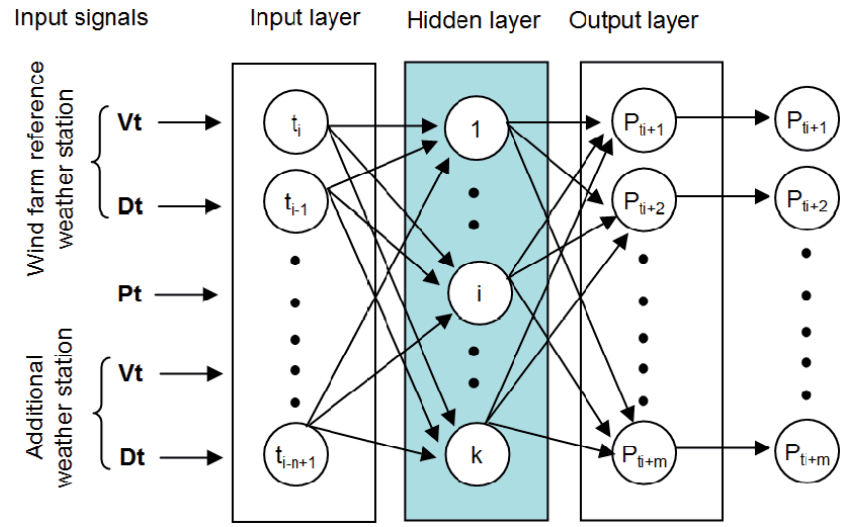


Figure 3.
Schematic representation of neural network for generation of forecasting models in case B.

Code	Height (magl)	Latitude (north)	Longitude (west)	Altitude (m)
WS1	40	27°54′08”	15°23′17”	16
WS2	10	27°51′36”	15°23′13”	3
WS3	10	28°27′10”	13°51′54”	24
WS4	10	28°57′07”	13°36′00”	10
WS5	13	28°01′36”	15°23′16”	5
WS6	10	28°07′30”	15°40′37”	472
WS7	10	27°56′08”	15°25′24”	186
WS8	10	28°02′35”	16°34′16”	51
WS9	40	29°05′47”	13°30′21”	457

Table 1.
Weather stations used in study.

the data of the reference WS of each wind farm (WS1 and WS9) are combined with the data of each of the seven other weather stations, WS-2 to WS-8 (**Table 1**). Therefore, for Case B, 175 different models are generated (25 × 7). After applying these models to each wind farm, their results are then compared.

The number of neurons in the input layer also varies, depending on the value of n, from 15 (n = 3) to 180 (n = 36).

The variation in the number of output layer neurons is the same as in Case A.

2.3 Metrics used to compare the different models

To compare the performance of the different models generated for Cases A and B, metrics (1) and (2) were used:

$$MARE = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{(T-r)}\sum_{i=1}^{T-r}\left(\frac{\left|P_j - \hat{P}_j\right|}{P_j}\right)_i = \frac{1}{m}MARE_j \tag{1}$$

$$R = \frac{1}{m}\sum_{j=1}^{m}\frac{\sum_{i=1}^{T-r}\left(P_{j_i} - \overline{P_j}\right)\times\left(\hat{P}_{ji} - \overline{\hat{P}_j}\right)}{\sqrt{\left[\sum_{i=1}^{T-r}\left(P_{ji} - \overline{P_j}\right)^2\right]\times\left[\sum_{i=1}^{T-r}\left(\hat{P}_{ji} - \overline{\hat{P}_j}\right)^2\right]}} = \frac{1}{m}\sum_{j=1}^{m}R_j \tag{2}$$

where: MARE is the mean absolute relative error for the forecast horizon; T is the number of data in the test stage (see **Figure 1**); $r = T\text{-}m\text{-}n$; $MARE_j$ is the mean absolute relative error for the forecasting period j; P_j and $\hat{P}_j$ are the actual and estimated wind farm power output in the forecasting period j, respectively; R is the mean value of Pearson's coefficient of correlation between the estimated and actual wind farm power output for the forecast horizon; and R_j is the mean Pearson correlation coefficient between the estimated and actual wind farm power output values for the forecasting period j.

The combined use of the two previous metrics is considered sufficient for the evaluation of the performance of the models and they have been widely used [38–41]. Alternatively, for the evaluation of future models, combinations of other metrics could be used [42]. For example, a combination of the Normalized Mean Absolute Error (NMAE) and the Index of Agreement (IoA) could be used.

3. Materials

The meteorological data (wind speed and direction) recorded by nine WSs located in four of the seven islands of the Canary Archipelago (**Table 1**) are used in this study. The mean hourly wind speed and direction data from 2008 are used in all cases. The heights of the WSs are expressed in metres above ground level (magl).

To validate and compare the results obtained with the different models, information corresponding to two wind farms (WF) located on two of the seven islands of the Canary Archipelago is used. **Tables 2** and **3** shows the geographic coordinates of the wind turbines (WT) of the two wind farms (WF1 and WF2). The hourly wind farm power output data for 2008 are used for this study.

Stations WS1 and WS9 (**Table 1**) are the reference weather stations of wind farms WF1 and WF2, respectively. The WS1 and WS9 data and the wind power production values are provided by the respective owners of the wind farms. The data from the seven additional WSs used in the study are provided by the Canary Islands Technological Institute (Spanish initials: ITC), a publicly owned R&D company run by the Regional Government of the Canary Islands and Spain's State Meteorological Agency (Spanish initials: AEMET).

Table 4 shows the results obtained for the coefficients of linear correlation (3) between the mean hourly wind speeds of the different WSs.

$$CC = \frac{\sum_{i=1}^{NG}\left(V_i - \overline{V}\right)\times\left(V_i^{'} - \overline{V'}\right)}{\sqrt{\sum_{i=1}^{NG}\left(V_i - \overline{V}\right)^2}\times\sqrt{\sum_{i=1}^{NG}\left(V_i^{'} - \overline{V'}\right)^2}} \tag{3}$$

where *CC* is the Pearson's coefficient of correlation between the wind speeds of two WSs; *NG* is the total number of data of the series. In this case, as a series of hourly data equivalent to one year is available, *NG* is equal to 8760. V_{i} and $V_i^{'}$ are the speeds at instant *i* of the two WSs subject to correlation; $\overline{V}$ and $\overline{V'}$ are the mean wind speeds of the two WSs subject to correlation for the available data series.

Code	*x* (m)	*y* (m)	*z* (m)
WF1-WT1	461764	3086314	3
WF1-WT2	461839	3086301	1
WF1-WT3	461681	3086067	5
WF1-WT4	461753	3086038	2

Table 2.
Geographic coordinates of wind turbines in WF1.

Code	x (m)	y (m)	z (m)
WF2-WT1	645043	3219819	486
WF2-WT2	645147	3219752	478
WF2-WT3	645186	3219638	473
WF2-WT4	645264	3219548	464
WF2-WT5	645333	3219462	456
WF2-WT6	645403	3219369	448
WF2-WT7	645406	3219213	440
WF2-WT8	645554	3219194	425
WF2-WT9	645664	3219133	405

Table 3.
Geographic coordinates of wind turbines in WF2.

No.	Coefficient of linear correlation								
	WS1	**WS2**	**WS3**	**WS4**	**WS5**	**WS6**	**WS7**	**WS8**	**WS9**
WS1	1.00	0.84	0.27	0.34	0.74	0.73	0.77	0.50	0.51
WS2	0.81	1.00	0.19	0.25	0.79	0.74	0.87	0.44	0.54
WS3	0.27	0.19	1.00	0.70	0.16	0.16	0.18	0.16	0.11
WS4	0.34	0.25	0.70	1.00	0.20	0.21	0.22	0.20	0.11
WS5	0.74	0.79	0.16	0.20	1.00	0.49	0.78	0.21	0.44
WS6	0.73	0.74	0.16	0.21	0.49	1.00	0.61	0.62	0.54
WS7	0.77	0.87	0.18	0.22	0.78	0.61	1.00	0.39	0.46
WS8	0.50	0.44	0.16	0.20	0.21	0.62	0.39	1.00	0.35
WS9	0.51	0.54	0.11	0.11	0.44	0.54	0.46	0.35	1.00

Table 4.
Coefficient of linear correlation between wind speeds of different weather stations in 2008.

4. Results and discussion

The discussion of the results centres on the two cases proposed in the methodology. For the various figures corresponding to the results, t-3 indicates that 2 periods prior to the forecasting period are chosen in addition to the forecasting period (t_i, t_i-1, t_i-2), and t + 3 indicates a forecasting horizon of 3 periods, t_i + 1, t_i + 2, t_i + 3, starting from the period for which the forecasting is being made, and so on for all combinations.

4.1 Discussion of results for case A

Figures 4 and 5 show the results for the MARE and R metrics for the 25 generated models. In practically all cases, the MARE and R values improve as n increases. The only exception is for case t-36 in comparison with t-24, where the improvement is minimal or not observed. In addition, the degree of improvement increases as m increases (t + 12, t + 24 and t + 36).

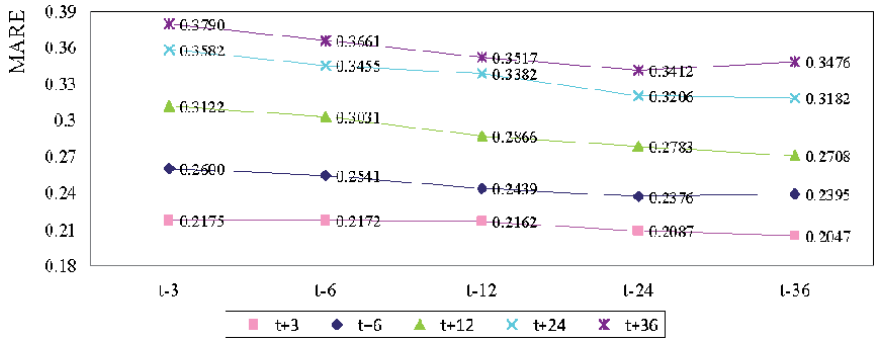


Figure 4.
MARE results in case A.

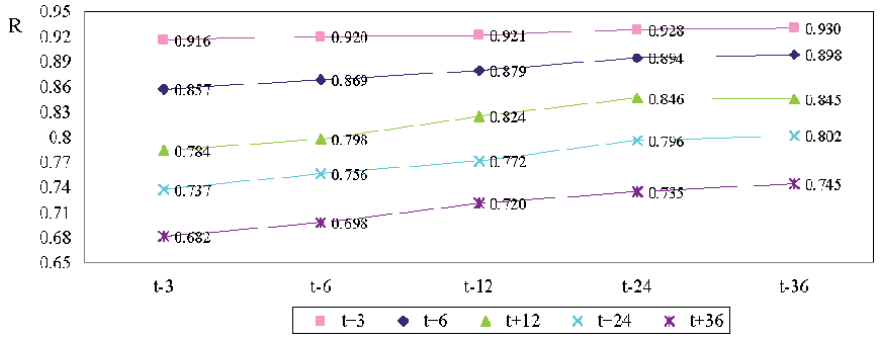


Figure 5.
R results in case A.

For the forecasting horizons $t + 12$, $t + 24$, $t + 36$, the maximum improvements obtained for MARE between the values for $n = 3$ and $n = 36$, are 13.3%, 11.2% and 10%, respectively. For the same cases but for R, the corresponding improvements are 7.9%, 8.9% and 9.2%, respectively.

To study the forecasting stability, an analysis has been made of the specific case of forecasting horizon $t + 24$, in which the number of periods to forecast is significant. **Figure 6** shows, for this specific case and differentiated according to n, the results of the variation of the relative error in the different forecasting periods, $MARE_j$. It can be seen how the relative error stabilizes earlier as n increases.

The forecasting stability is analyzed for all the forecasting horizons (**Figure 7**). This analysis is made on the basis of the standard deviation of relative error in the forecasting horizon:.

$$SDV = \sqrt{\frac{\sum_{j=1}^{m}\left(MARE_j - MARE\right)^2}{m-1}} \tag{4}$$

where SDV is the mean standard deviation of the MARE for a forecasting time horizon m.

It can be seen in **Figure 7** that for all the forecasting horizons, the SDV/MARE value decreases significantly as the number of prior hours n increases. This significant improvement in the stability of models is observed even for the lowest forecasting horizons. Only for the particular case of forecasting horizon $t + 3$ and when the horizon passes from t-24 to t-36, no improvement is observed.

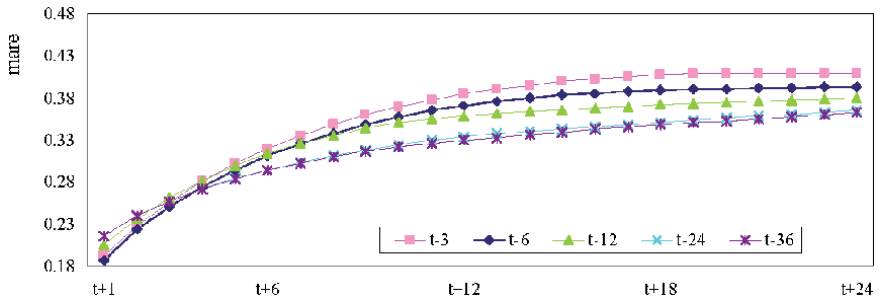


Figure 6.
MARE variation of different prediction periods: Case of a forecasting horizon t + 24.

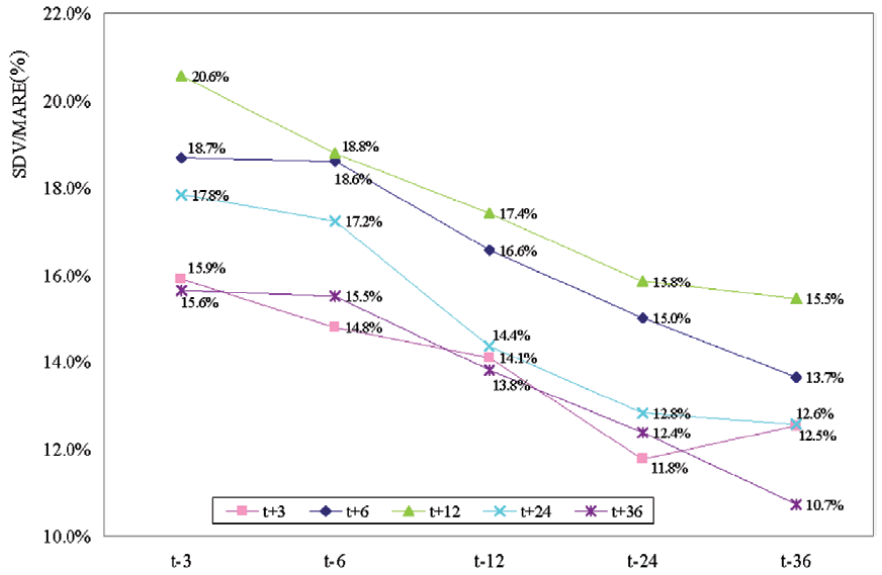


Figure 7.
Stability of relative error SDV in forecasting horizon.

By way of example, we will now proceed to analyze the specific cases of the forecasting models $t + 12$ and $t + 24$. To date, in the ANN models studied in the literature, the number of prior periods n chosen to generate the models has always been fixed. Assume that the n chosen for a standard model is 12. In this case, the MARE value is 0.2866 for the $t + 12$ model and 0.3382 for the $t + 24$ model (**Figure 4**). The corresponding values for the stability of the relative error are 17.4% and 14.4% (**Figure 7**), respectively. According to the analysis made with Case A, the performance of these models can be improved by choosing a higher value of n. If n is 24, the MARE values decrease to 0.2783 and 0.3206, respectively (**Figure 4**). Similarly, for an n of 24, the stability of the relative error in the forecasting improves to the values of 15.8% and 12.8%, respectively (**Figure 7**).

4.2 Discussion of results for case B

For the analysis of Case B, the MARE and R results of this case, with two WSs, are compared with those of Case A, with one WS. For this purpose, (5) and (6) are used.

$$\Delta MARE = \frac{1}{7}\sum_{p=1}^{7}\frac{MARE_{p_{(with\,2\,WSs)}} - MARE_{(with\ 1WS)}}{MARE_{(with\,1WS)}} \times 100\% \quad (5)$$

$$\Delta R = \frac{1}{7}\sum_{p=1}^{7}\frac{R_{p_{(with\,2\,WSs)}} - R_{(with\,1WS)}}{R_{(with\,1WS)}} \times 100\% \quad (6)$$

It can be seen in **Figures 8** and **9** how all the models generated for Case B obtain an additional improvement in performance to that already obtained for Case A. This additional improvement is in relation to ANN models developed to date which always use exclusive data from a single WS.

It can also be observed that, in general, the degree of improvement increases as m increases. This degree of improvement slows down for forecasting horizons longer than 24 hours.

The maximum additional improvements in model performance are seen in forecasting horizons $t + 24$ and $t + 36$ (7.5% and 5.5% for MARE and 3.7% and 5.4% for R, respectively). Even for the shortest forecasting horizons, $t + 3$ and $t + 6$, the maximum improvements in the MARE metric are significant (3% and 4.9%, respectively).

Continuing with the specific example proposed in the analysis of results for Case A (using models $t + 12$ and $t + 24$), **Figure 10** shows the additional improvements in performance that can be obtained through the incorporation in the input layer of data from a second WS (Case B).

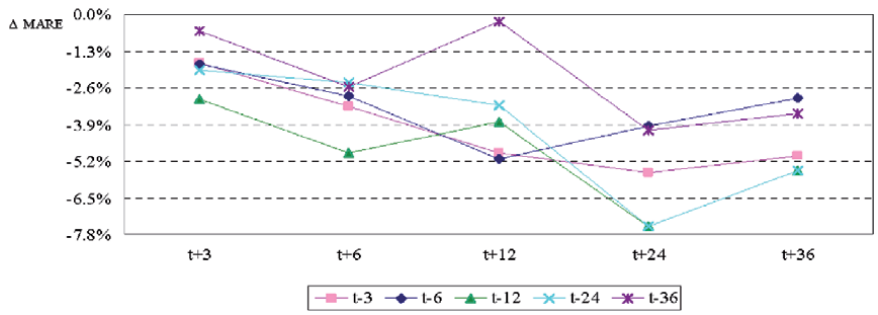


Figure 8.
Comparison of MARE results for cases A and B.

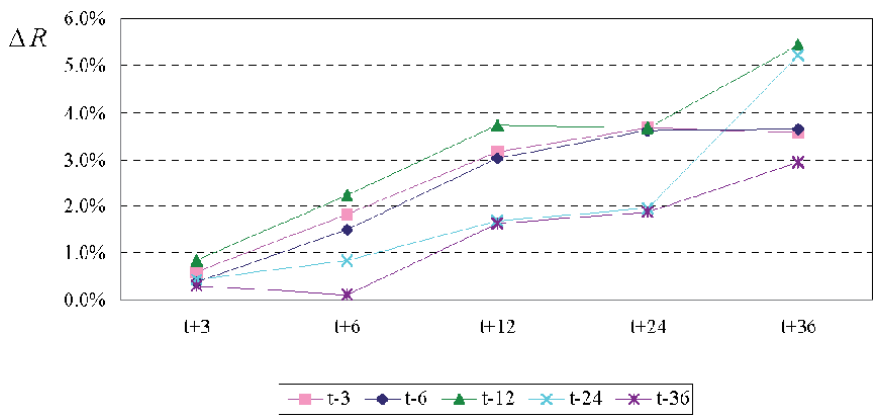


Figure 9.
Comparison of results obtained for R for cases A and B.

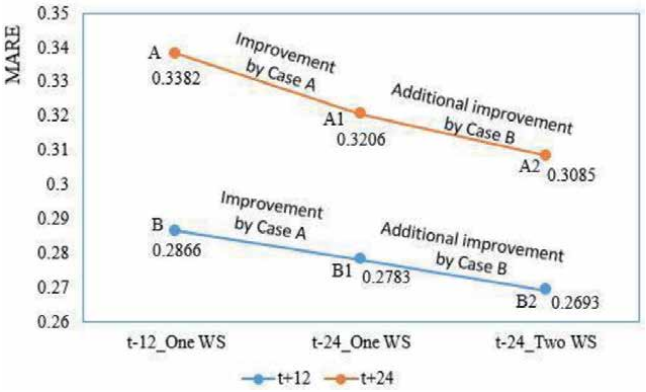


Figure 10.
Improvements in error for two specific models due to implementation of cases A and B.

Points A and B represent the error obtained when using a fixed n of 12 and only data from the reference WS of the wind farm. Points A1 and B1 represent the improvements obtained in the error when n is increased to 24. Points A2 and B2 represent the additional improvements obtained in the error when, in Case B, the data from a second WS are incorporated in the input layer of the ANN. For the two specific examples given, the overall improvements obtained by combining Cases A and B amount to 8.78% and 6.04%, respectively.

5. Conclusion

A series of interesting conclusions can be drawn from the results of this study with respect to possible improvements in the performance of ANN models for the short-term forecasting of wind farm power output.

The performance of the new ANN models generated for each forecast horizon improves with the increase in the number of prior 1-h periods (periods prior to the prediction hour), n, chosen for incorporation of the input layer parameters. For the forecasting horizons $t + 12$, $t + 24$ and $t + 36$, the maximum improvements obtained for MARE are 13.3%, 11.2% and 10%, respectively; and for R, the corresponding improvements are 7.9%, 8.9% and 9.2%, respectively.

A study is also made of the stability of the mean relative error for the different forecasting periods and for each forecasting horizon m. As n increases the stability of the error in the forecasting improves significantly for all forecasting horizons.

Additionally, in all the new models generated, the incorporation in the input layer of ANN of meteorological data from a second WS also improves the performance of the traditional models generated exclusively with data from the reference station of the wind farm. In general terms, the degree of improvement in model performance increases with m, attaining improvements in the MARE and R of up to 7.5% and 5.4%, respectively.

In view of the conclusions drawn from the present study, the original contributions described in this manuscript could be implemented in existing ANN models to optimize their results.

Acknowledgements

This research has been co-funded by ERDF funds, INTERREG MAC 2014-2020 programme, within the ENERMAC project (MAC/1.1a/117). No funding sources

had any influence on study design, collection, analysis, or interpretation of data, manuscript preparation, or the decision to submit for publication.

Conflict of interest

The authors declare no conflict of interest.

Author details

Sergio Velázquez-Medina[1*] and Ulises Portero-Ajenjo[2]

1 Department of Electronics and Automatics Engineering, Universidad de Las Palmas de Gran Canaria, Canary Islands, Spain

2 School of Industrial and Civil Engineering, University of Las Palmas de Gran Canaria, Canary Islands, Spain

*Address all correspondence to: sergio.velazquezmedina@ulpgc.es

References

[1] C. G. Justus, K. Mani and A.S. Mikhail, "Interannual and month-to-month variations of wind speed", Journal of Applied Meteorology, vol 18, no. 7, pp. 913-920, Jul. 1979.

[2] R. Baker, S. N. Walker, J.E. Wade, "Annual and seasonal variations in mean wind speed and wind turbine energy production" Solar Energy, vol. 45, no. 5, pp. 285-289, 1990.

[3] K. Klink, "Trends and interannual variability of wind speed distributions in Minnesota", Journal of Climate, vol. 15, no. 22, pp. 3311-3317, 2002.

[4] Wind Energy Handbook, 2nd edition, John Wiley & Sons, 2011

[5] L. Landberg, L. Myllerup and O. Rathmann, et al. "Wind resource estimation—An overview", Wind Energy, vol. 6, no. 3, pp. 261-271, Jul. 2003.

[6] A. Aziz, A. M. Than and A. Stojcevski (2018, Jul). Issues and mitigations of wind energy penetrated network: Australian network case study. Journal of Modern Power System and Clean Energy. [Online]. Available: https://doi.org/10.1007/s40565-018-0430-4

[7] A. Basit, A. D. Hansen and P. E. Sørensen et al. (2015/Nov.). Real-time impact of power balancing on power system operation with large scale integration of wind power. Journal of Modern Power System and Clean Energy. [Online]. 5(2), pp. 202-210. Available: https://link.springer.com/article/10.1007/s40565-015-0163-6

[8] T. Mahmoud, Z. Y. Dong and J. Ma, "Advanced method for short-term wind power prediction with multiple observation points using extreme learning machines", *The Journal of Engineering*, vol. 2018, no. 1, pp. 29.38, Mar. 2018.

[9] P. Du, H. Hui and N. Lu, "Procurement of regulation services for a grid with high-penetration wind generation resources: a case study of ERCOT", IET Generation, Transmission and Distribution, vol. 10, no. 16, pp. 4085-4093, 2016.

[10] A. Basit, A. D. Hansen and M. Altin et al. (2016/Jul.). Compensating active power imbalances in power system with large-scale wind power penetration. Journal of Modern Power System and Clean Energy. [online]. 4(2), pp. 229-237Available:https://link.springer.com/article/10.1007/s40565-015-0135-x

[11] O. Abedinia and N. Amjady, "Net demand prediction for power systems by a new neural network-based forecasting engine", Complexity, vol 21, pp. 296-308, Jul. 2016.

[12] S. Sreekumar, K. Chand Sharma, R. Bhakar. "Grey System Theory Based Net Load Forecasting for High Renewable Penetrated Power Systems". Technology and Economics of Smart Grids and Sustainable Energy, https://doi.org/10.1007/s40866-020-00094-4, 2020

[13] M. Bagheri, O. Abedinia and M. Salary et al. "Direct and indirect prediction of net demand in power systems based on syntactic forecast engine". IEEE International Conference on Environment and Electrical Engineering, Palermo, Italy, Jun. 2018.

[14] Y. Jiang, X. Chen and K. Yu1 et al. (2017). Short-term wind power forecasting using hybrid method based on enhanced boosting algorithm. *Journal of Modern Power System and Clean Energy*. [online]. 5(1), pp. 126-133. Available: https://link.springer.com/article/10.1007/s40565-015-0171-6

[15] H. Chen, F. Li and Y. Wang. (2016/Sep.). Wind power forecasting based on outlier smooth transition autoregressive

GARCH model. Journal of Modern Power System and Clean Energy. [online]. 6(3), pp. 532-539. Available: https://doi.org/10.1007/s40565-016-0226-3

[16] M. Xu, Z. Lu and Y. Qiao et al. (2017/Jan.). Modelling of wind power forecasting errors based on kernel recursive least-squares method. Journal of Modern Power System and Clean Energy. [online]. 5(5), pp. 735-745. Available: https://link.springer.com/article/10.1007/s40565-016-0259-7

[17] D. Kim and J. Hur, "Short-term probabilistic forecasting of wind energy resources using the enhanced ensemble method", Energy, vol. 157, pp. 211-226, 2018.

[18] N. Huang , E. Xing and G. Cai et el. "Short-term wind speed forecasting based on low redundancy feature selection", *Energies*, vol. 11, no. 7, 1638, Jul. 2018.

[19] Z. Fei, L. Peng-Cheng, G. Lu, L. Yong-Qian, R. Xiao-Ying. "Application of autoregressive dynamic adaptive (ARDA) model in realtime wind power forecasting". Renewable Energy, vol. 169, 129e143, 2021

[20] P. Kumar Singh, N. Singh, R. Negi. "Short-Term Wind Power Prediction Using Hybrid Auto Regressive Integrated Moving Average Model and Dynamic Particle Swarm Optimization". International Journal of Cognitive Informatics and Natural Intelligence, vol. 15, • Issue 2, 2020

[21] S. J. Ghoushchi, S. Manjili, A. Mardani, M. K. Saraji. "An extended new approach for forecasting short-term wind power using modified fuzzy wavelet neural network: A case study in wind power plant". Energy, vol. 223, 120052, 2021

[22] Y. Rui, L. Dengxuan, W. Yifeng, C. Weidong. "Forecasting method of monthly wind power generation based on climate model and long short-term memory neural network". Global Energy Interconnection, vol. 3, n° 6, 2020

[23] T. Liu, S. Liu and J. Heng et al. "A new hybrid approach for wind speed forecasting applying Support Vector Machine with ensemble empirical mode decomposition and Cuckoo Search Algorithm", Applied Sciences (Switzerland), vol. 8, no. 10, pp. 1754, Oct. 2018.

[24] O. Abedinia, D. Raisz and N. Amjady, "Effective prediction model for Hungarian small-scale solar power output", IET Renewable. Power Generation, vol. 11, no. 13, pp. 1648-1658, 2017.

[25] Y. Zhang, K. Liu and L. Qin et al. "Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods", Energy Conversion and Management, 112, pp. 208-219, Jan. 2016.

[26] A. Zameer, J. Arshad and A. Khan et al. "Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks", Energy Conversion and Management, 134, pp. 361-372, 2017.

[27] M. Felder, F. Sehnke and K. Ohnmeiß, et al. (2018/Jul.). Probabilistic short term wind power forecasts using deep neural networks with discrete target classes. Advances in Geosciences. [online]. 45, pp. 13-17. Available: https://doi.org/10.5194/adgeo-45-13-2018

[28] Najeebullah, A. Zameer and A. Khan et al. "Machine Learning based short term wind power prediction using a hybrid learning model", Computers and Electrical Engineering, vol. 45, pp. 122-133, 2015.

[29] M. Morina, F. Grimaccia and S. Leva et al. "Hybrid weather-based ANN for forecasting the production of a real

wind power plant", Proceedings of the International Joint Conference on Neural Networks, 7727858, pp. 4999-5005, Oct. 2016.

[30] P. Mandal, H. Zareipour and W. D. Rosehart, "Forecasting aggregated wind power production of multiple wind farms using hybrid wavelet-PSO-NNs", International Journal of Energy Research, vol. 38, no. 13, pp. 1654-1666, Feb. 2014.

[31] G. Zhang, L. Zhang and T. Xie, "Prediction of short-term wind power in wind power plant based on BP-ANN", Proceedings IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, Xi'an, China, pp. 75-79, Oct. 2016

[32] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power", Renewable and Sustainable Energy Reviews, vol. 34, pp. 243-254, Jun. 2014.

[33] D. Lee and R. Baldick, "Short-term wind power ensemble prediction based on Gaussian processes and neural networks", *IEEE Transactions on Smart Grid*, vol. 5, no. 1, 6606922, pp. 501-510, Jan. 2014.

[34] N. Amjady and O. Abedinia, "Short term wind power prediction based on improved Kriging interpolation, empirical mode decomposition, and closed-loop forecasting engine", Sustainability (Switzerland), vol. 9, no. 11, pp. 2104, Nov. 2017.

[35] J. C. Principe, N. R. Euliano and W. C. Lefebvre, "Neural and Adaptive Systems. Fundamentals Through Simulations", 1st ed. New York: John Wiley & Sons, Inc., 2000.

[36] T. Masters,"Practical Neural Network Recipes in C++", 1st ed. California: Morgan Kaufmann Publishers, 1993.

[37] N. R. Draper and H. Smith, "Applied regression analysis", 3rd ed. John Wiley & Sons, Inc, Apr. 1998.

[38] S. Velázquez-Medina, J.A. Carta, U. Portero-Ajenjo. "Performance sensitivity of a wind farm power curve model to different signals of the input layer of ANNs: Case studies in the Canary Islands". Complexity, vol. 2019, article number 2869149, 11 pages

[39] W. Jujie, L. Yaning. "Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy". Applied Energy, vol. 230, pp. 429-443, 2018

[40] S. Díaz, J. A. Carta, J. M. Matías. "Performance assessment of five MCP models proposed for the estimation of long-term wind turbine power outputs at a target site using three machine learning techniques". Applied Energy, vol. 209, pp. 455-477, 2018

[41] S. Díaz, J. A. Carta, J. M. Matías. "Performance assessment of five MCP models proposed for the estimation of long-term wind turbine power outputs at a target site using three machine learning techniques". Applied Energy, vol. 209, pp. 455-477, 2018

[42] J.M. González-Sopeña, V. Pakrashi, B. Ghosh. "An overview of performance evaluation metrics for short-term statistical wind power forecasting". Renewable and Sustainable Energy Reviews, vol. 138, 110515, 2021.

Edited by Ricardo López-Ruiz

Over two parts, this book examines the meaning of complexity in the context of systems both social and natural. Chapters cover such topics as the traveling salesman problem, models of opinion dynamics creation, a universal theory for knowledge formation in children, the evaluation of landscape organization and dynamics through information entropy indicators, and studying the performance of wind farms using artificial neural networks. We hope that this book will be useful to an audience interested in the different problems and approaches that are used within the theory of complexity

Published in London, UK

IntechOpen

ISBN 978-1-83962-855-9