Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Electronic Notes in Theoretical Computer Science 264 (2010) 105–123
Electronic Notes in Theoretical Computer Science
www.elsevier.com/locate/entcs

Applications of Algebra and Coalgebra in Scientific Modelling

Illustrated with the Logistic Map

Michael Hauhs[1] Baltasar Trancón y Widemann[2]

Ecological Modelling
University of Bayreuth
Bayreuth, Germany

Abstract

In computer science, the algebra–coalgebra duality serves as a formal framework for connecting the perspectives of state-based and behavior-based models. In other sciences such as ecology, these perspectives are seemingly harder to reconcile. We explore modelling paradigms, in the sense of philosophy of science, as an intermediate step in translating the (co)algebraic framework from computer science into applications in ecology. We illustrate the application potential of this approach with a simple model from theoretical ecology: the logistic map. Several versions of algebraic models with progressively more sophisticated carriers and operations are introduced and finally contrasted with a corresponding coalgebraic model. We illustrate two modelling paradigms with these examples. Only one of these has traditionally been used in ecology. The second one, which is based on a coalgebraic dualisation, offers new modelling perspectives in ecology and environmental science.

Keywords: algebra, coalgebra, state, behavior, model, paradigm, scientific method, dynamic system, ecology, logistic map

1 Introduction

Scientific modelling, the task of relating theories and data, is a multi-faceted problem without a single universal solution. Besides the particular discipline of science under study, it is necessarily connected to the polar areas of philosophy and mathematics.

A fundamental dichotomy from both the philosophic and the mathematical viewpoint is the choice between *state* and *behavior* as the primary ontological category of system properties. There are some scientific disciplines where one is clearly dominant: Physical sciences tend to be state-based, whereas social sciences tend to be

[1] Email: Michael.Hauhs@uni-bayreuth.de
[2] Email: Baltasar.Trancon@uni-bayreuth.de

1571-0661/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.entcs.2010.07.016

behavior-based. But there is also a middle ground covered by life-related sciences, in particular ecology as the science of living systems in an open environment. These sciences pose especially interesting and hard challenges to the modeller, because neither state nor behavior alone seem to be sufficient for comprehensive system descriptions.

In most scientific fields, the primacy of either state or behavior is correlated with the degree of formalization: State-based models tend to be given in mathematical formulae, whereas behavior-based models tend to be given in narrative prose. Computer science is rather distinguished by the fact that it provides methods to render both perspectives with comparable formal rigor, and to unify them in common frameworks. Of these frameworks, we regard the duality of universal algebra and coalgebra as particularly promising for scientific modelling, for several reasons:

(i) There are vast bodies of theoretical results on how to apply algebra and coalgebra to state-based (e.g. [5]) and behavior-based (e.g. [9]) system models, respectively.

(ii) The duality is a precise relationship within the meta-framework of category theory, as opposed to a mere philosophical complementarity [16].

(iii) The usefulness of commuting diagrams similar to those underlying the categorial formulation of (co)algebra for theoretical biology has already been established [15].

Our present work should be understood as a small step towards leveraging the tools of theoretical computer science for theoretical ecology. This overall goal is not easy to achieve; not least because the structural mathematics of computer science remain obscure and inaccessible to the more classically trained ecologist. As an intermediate, more modest goal, we aim at extending the repertoire of scientific ecological modelling with methods originally designed for the description of systems of logic, control and computation. Towards this end, we shall presently discuss a system that is simple and idealized, yet of some popularity in theoretical ecology. We shall illustrate that modelling questions concerning this system fall into the two aforementioned dual categories, and how they can be mapped to algebraic and coalgebraic formulations, respectively. Our focus here shall be the systematic development of modelling techniques from basic universal (co)algebra and their interpretation from the meta-viewpoint of philosophy of science; the connection to more realistic and practical ecological problems is outside the scope of this article.

1.1 Scientific Modelling

Our modelling examples will be idealized. We use the notions proposed by [6] in which modelling is composed of two steps: The first step replaces a real-world phenomenon, the *target system*, with an idealised system described in words, the *model system* (see Figure 1 ibid.). In the historical case of astronomical models of planets, the physical objects were replaced by idealised, homogenous spheres with point mass. Here we use a population of organisms and its temporal variation

by growth as the target system and replace it with a spatially homogenous model system: The model system is then in the second step described by the logistic map.

A modelling paradigm, as introduced by Kuhn, links an aspect of the empirical world, the model system (in the sense above), with mathematics. It has to include a recipe, how to fill/relate the description of the word model with data on the one hand and how to symbolise the description and apply mathematics on the other hand.

Several options for this task exist. They are vastly different with respect to their reputation in science, to the extent that sometimes one paradigm, the physical one, is identified with the scientific method as such. However, empirics and management practice, at least for ecological problems, appear determined to remain methodologically diverse. We do not take sides in this dispute, and present two dual modelling paradigms without judging their relative applicability a priori.

Each of the two paradigms emphasizes one of the two ontological categories: The *functional* paradigm is based on observable states; behavior is a secondary notion that arises of the change of states under a dynamic law. The *interactive* paradigm focuses directly on behavior; state arises from the history of choices of agents. The latter paradigm is uncommon in most "hard" sciences. Again, computer science is an exception; cf. the famous Turing test.

It is here where we expect the impact of coalgebra. The new theoretical approach may formalise a model paradigm which is already implicitly used in ecological practice, but which has not been recognised in theoretical ecology [17]. A corresponding problem in the philosophy of science is the epistemological classification of computer simulations [8].

Using coalgebra to model natural phenomena is not (yet) a popular approach. This is no surprise, because few natural scientists are even aware of the existence of such a theory. The gap between the research programs of natural sciences on one hand and of theoretical computer science on the other hand makes it difficult to exchange abstract notions and theoretical frameworks. Rigorous study of scientific modelling and its integration of "mindset" and "toolkit" can be beneficial to mutual understanding.

The logistic map has been chosen as an objct of study for its simplicity, not for its immediate practical relevance. For a more relevant example of ecological behavior, consider the idealised case of a domesticated species in which evolutionary change can be supressed deliberately. The complete space of possible behavior *under human management* can then be derived from its documented growth history. The scientific task is to comprehensively represent patterns of this history along with proper goals and intervention norms, in order to allow a sustained continuation of the past behavior, but without being able to reconstruct the system after irreversible failure, such as extinction of the species.

1.2 The Logistic Map

The so-called logistic map [10] is related to the logistic equation published by Verhulst in 1838, which was one of founding concepts of theoretical population biol-

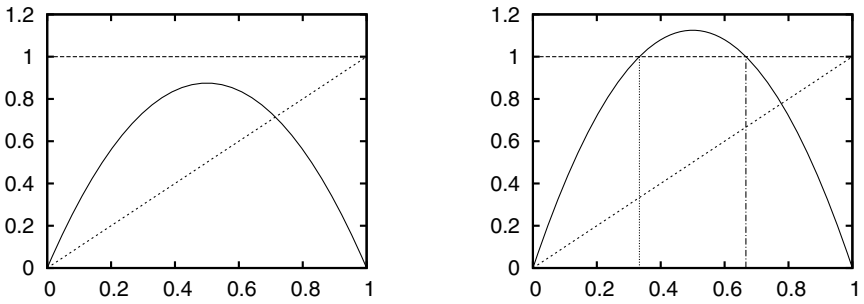


Fig. 1. The logistic map f_r for $r < 4$ (left) and $r > 4$ (right)

ogy [1]. It has been criticised for being oversimplified, but is still a reference concept for more realistic models. It is used as the introductory example in a standard textbook on theoretical biology [12].

Definition 1.1 For a real parameter $r > 0$, the **logistic map** is defined as the real function

$$f_r(x) = rx(1 - x)$$

restricted in both domain and range to the unit interval $\mathbb{I} = [0, 1]$; see Figure 1.

- For $r \leq 4$, the function f_r is totally defined on $\mathbb{I}$.
- For $r > 4$, the function f_r is only partially defined on $\mathbb{I}$: $f_r(x) \notin \mathbb{I}$ for some $x \in \mathbb{I}$.

The single parameter r is interpreted as the effective growth rate of the system. The state of the system is interpreted as population density, normalised by the carrying capacity of the system with respect to the given environment. Any relation with the environment is encoded into the carrying capacity parameter, hence this gives rise to a discrete autonomous dynamic system.

Dynamic systems with the state space $\mathbb{I}$ and the step function f_r exhibit a variety of interesting modes of behavior, depending on the value of r: from certain extinction through stable fixed points and periodic solutions of all periods to deterministic chaos with strange attractors.

The logistic map has been investigated with the methods of symbolic dynamics as an important case of a complex, chaotic system. In this role it has also been used as an application of coalgebra [16]. This, to our knowledge, has been the first connection between coalgebra and models used in biology. Here we use the well-known features of this map for reviewing the various roles in which dynamic models can be used in ecological modelling.

Time is discrete in a system with the step function f_r. This is not necessarily an idealization for biological systems; e.g. generation times. But the state space $\mathbb{I}$ is idealized as continuous. For the application of symbolic dynamics and to accomodate the realistic assumption that measurements cannot be made arbitrarily precise, we discretize observations as partitions of the state space, specified by the assignment

of symbols from a finite alphabet. It suffices to consider the most coarse-grained case.

Definition 1.2 The **binary unit partition** is defined as a function $c : \mathbb{I} \to 2 = \{0, 1\}$

$$c(x) = \begin{cases} 0 & \text{if } x < \frac{1}{2} \\ 1 & \text{if } x \geq \frac{1}{2} \end{cases}$$

Note that f_r is not reversible (injective), but the tupling $\langle c, f_r \rangle : \mathbb{I} \to 2 \times \mathbb{I}$ is.

- For $r < 4$, the function $\langle c, f_r \rangle^{-1}$ is only partially defined on $2 \times \mathbb{I}$.
- For $r \geq 4$, the function $\langle c, f_r \rangle^{-1}$ is totally defined on $2 \times \mathbb{I}$.

This binary partition of the logistic dynamic system has been used in [4] to demonstrate that the apparent complexity of a system depends crucially on the viewpoint.

2 Formal Prerequisites

The mathematical structures underlying not only our example model system, but more or less directly every dynamic system, are the sequential data structures: finite and infinite sequences over a fixed set of elements. These structures and the usual ways of reasoning with them have well-understood representations in terms of algebra and coalgebra.

2.1 Strings and Streams

Definition 2.1 The set A^* is called the set of **finite sequences** or **strings** over A. It is generated by the free constructors $\mathrm{cons}_A : A \times A^* \to A^*$ and $\mathrm{nil}_A \in A^*$. The destructors are the unique partial functions $\mathrm{hd}_A : A^* \nrightarrow A$ and $\mathrm{tl}_A : A^* \nrightarrow A^*$ such that

$$\mathrm{hd}_A\big(\mathrm{cons}_A(a, w)\big) = a \qquad \mathrm{hd}_A(\mathrm{nil}_A) \text{ undefined}$$
$$\mathrm{tl}_A\big(\mathrm{cons}_A(a, w)\big) = w \qquad \mathrm{tl}_A(\mathrm{nil}_A) \text{ undefined}$$

- We omit all subscript annotations where no ambiguity arises.
- We define the subsets $A^n \subset A^*$ of strings of a fixed length n inductively as

$$A^{n+1} = \mathcal{P}(\mathrm{cons}_A)(A \times A^n) \qquad A^0 = \{\mathrm{nil}_A\}$$

Then

$$A^* = \bigcup_{n=0}^{\infty} A^n \qquad A^+ = \bigcup_{n=1}^{\infty} A^n = A^* \setminus \{\mathrm{nil}_A\}$$

- We informally write $a_1 \ldots a_n$ for $\mathrm{cons}(a_1, \ldots, \mathrm{cons}(a_n, \mathrm{nil}) \ldots)$.
- In particular, we abbreviate a singleton string $\mathrm{cons}(a, \mathrm{nil})$ to a.

- For any function $f : A \to B$, we write $f^* : A^* \to B^*$ for the elementwise mapping

$$f^*\big(\mathrm{cons}_A(a, w)\big) = \mathrm{cons}_B\big(f(a), f^*(w)\big) \qquad f^*(\mathrm{nil}_A) = \mathrm{nil}_B$$

  This turns * into a functor.

The choice of cons and nil as the constructors of strings suggest that the organization of data in a string obey the *stack* principle: data elements are accumulated and removed at the left end of a string only. The following auxiliary function handles a special case of this principle, namely the accumulation of data arising from the iterated application of a given function.

Definition 2.2 Let A be any set and $f : A \nrightarrow A$ a partial function. The partial function $\mathrm{push}(f) : A^* \nrightarrow A^*$ is defined as

$$\mathrm{push}(f)(w) = \mathrm{cons}\Big(f\big(\mathrm{hd}(w)\big), w\Big)$$

- *Strict* application is implied: $\mathrm{push}(f)$ is undefined at w if $\mathrm{hd}(w)$ or $f(\mathrm{hd}(w))$ is undefined. In particular, $\mathrm{push}(f)(\mathrm{nil})$ is never defined.
- Note that $\mathrm{push}(f)^n : A^m \to A^{m+n}$ for $m > 0$ and $n \geq 0$. In slight abuse of notation we define $\mathrm{push}(f)^{-n} : A^m \to A^{m-n}$ for $m \geq n \geq 0$ as the retraction

$$\mathrm{push}(f)^{-n} = \mathrm{tl}^n$$

Example 2.3 The expression $\mathrm{push}(\mathrm{succ})^n(0)$ yields a countdown from n.

Definition 2.4 The set $A^\omega = (\mathbb{N} \to A)$ is called the set of **infinite sequences** or **streams** over A. Its elements are of the form $\mathrm{cons}_A(a, s)$ for any $a \in A; s \in A^\omega$, with

$$\mathrm{cons}_A(a, s)(n) = \begin{cases} a & \text{if } n = 0 \\ s(n-1) & \text{if } n > 0 \end{cases}$$

We write

$$\mathrm{hd}_A(s) = s(0) \qquad \mathrm{tl}_A(s) = s \circ \mathrm{succ}$$

for the total destructors.

2.2 (Co)Algebras of Affine Type

Definition 2.5 The family of **affine** functors $\mathcal{A}^A_B : \mathbf{Set} \to \mathbf{Set}$ is defined as

$$\mathcal{A}^A_B(X) = A \times X + B \qquad \mathcal{A}^A_B(f) = \mathrm{id}_A \times f + \mathrm{id}_B$$

We write

$$\mathrm{go}_A : A \times X \to \mathcal{A}^A_B(X) \qquad \mathrm{stop}_B : B \to \mathcal{A}^A_B(X)$$

for the left and right injection, respectively.

Lemma 2.6 *Affine functors have initial algebras. The structure* $\big(A^* \times B, \alpha = [\alpha_1, \alpha_2]\big)$ *with*

$$\alpha_1\big(a,(w,b)\big) = \big(\mathrm{cons}(a,w),b\big) \qquad \alpha_2(b) = (\mathrm{nil},b)$$

is an initial $\mathcal{A}_B^A$*-algebra. The unique homomorphism or* ***catamorphism*** h *into any* $\mathcal{A}_B^A$*-algebra* $\big(C, \gamma = [\gamma_1,\gamma_2]\big)$ *is defined recursively as*

$$h\big(\mathrm{cons}(a,w),b\big) = \gamma_1\big(a,h(w,b)\big) \qquad h(\mathrm{nil},b) = \gamma_2(b)$$

Lemma 2.7 *Affine functors have final coalgebras. The structure* $\big((A^* \times B) + A^\omega, \phi = [\phi_1,\phi_2]\big)$ *with*

$$\phi_1\big(\mathrm{cons}(a,w),b\big) = \mathrm{go}\big(a,\iota_1(w,b)\big) \qquad \phi_2(s) = \mathrm{go}\big(\mathrm{hd}(s),\iota_2(\mathrm{tl}(s))\big)$$
$$\phi_1(\mathrm{nil},b) = \mathrm{stop}(b)$$

where ι_1,ι_2 *are the injections into* $(A^* \times B) + A^\omega$ *is a final* $\mathcal{A}_B^A$*-coalgebra. The unique homomorphism or* ***anamorphism*** h *from any* $\mathcal{A}_B^A$*-coalgebra* (C,γ) *is defined corecursively as*

$$\phi\big(h(c)\big) = \begin{cases} \mathrm{go}\big(a,h(c')\big) & \text{if } \gamma(c) = \mathrm{go}(a,c') \\ \mathrm{stop}(b) & \text{if } \gamma(c) = \mathrm{stop}(b) \end{cases}$$

Instantiating A or B with the empty set or the singleton set $1 = \{\star\}$ yields cases of special interest.

(i) The affine functor $\mathcal{A}_B^1$. The operations of $\mathcal{A}_B^1$-algebras are of type $\gamma : 1 \times C + B \to C$. They are in natural one-to-one correspondence to pairs (f,g) of type $f : C \to C$ and $g : B \to C$, namely $(f,g) \leftrightarrow [f \circ \pi_2, g]$. The carrier of the canonical initial $\mathcal{A}_B^1$-algebra (Lemma 2.6) simplifies to $\mathbb{N} \times B$ by reading 1^* as a unary number system. Its operation is specified by (f_0,g_0) with $f_0(n,b) = (n+1,b)$ and $g_0(b) = (0,b)$. The catamorphism i into the $\mathcal{A}_B^1$-algebra specified by (f,g) is the *iteration* operator

$$i(n,b) = f^n\big(g(b)\big)$$

(ii) The affine functor $\mathcal{A}_1^A$. The operations of $\mathcal{A}_1^A$-algebras are of type $\gamma : A \times C + 1 \to C$. They are in natural one-to-one correspondence to pairs (f,e) of type $f : A \times C \to C$ and $e \in C$, namely $(f,e) \leftrightarrow [f,\hat{e}]$ where $\hat{e}(*) = e$. The carrier of the canonical initial $\mathcal{A}_1^A$-algebra simplifies to A^*. Its operation is specified by $(\mathrm{cons}_A, \mathrm{nil}_A)$. The catamorphism j into the $\mathcal{A}_1^A$-algebra specified by (f,e) is the *fold* operator

$$j\big(\mathrm{cons}(a,w)\big) = f\big(a,j(w)\big) \qquad j(\mathrm{nil}) = e$$

(iii) The affine functor $\mathcal{A}_\varnothing^A$. The initial $\mathcal{A}_\varnothing^A$-algebra is empty. The operations of $\mathcal{A}_\varnothing^A$-coalgebras are of type $\gamma : C \to A \times C + \varnothing$. They are in natural one-to-one correspondence to pairs (h,t) of type $h : C \to A$ and $t : C \to C$, namely

$(h,t) \leftrightarrow \iota_1 \circ \langle h,t \rangle$. The carrier of the canonical final $\mathcal{A}^A_{\varnothing}$-coalgebra (Lemma 2.7) simplifies to A^ω. Its operation is specified by $(\mathrm{hd}_A, \mathrm{tl}_A)$. The anamorphism k from the $\mathcal{A}^A_{\varnothing}$-coalgebra specified by (h,t) is the *unfold* operator

$$k(c) = \mathrm{cons}\Big(h(c), k\big(t(c)\big)\Big)$$

(iv) The affine functor $\mathcal{A}^{\varnothing}_B$ is degenerate and equivalent to the constant functor B.

We shall demonstrate that each nondegenerate case corresponds to a scientific modelling scenario. We use the pair notation of the preceding paragraphs to specify operations, in order to avoid cluttering diagrams with uninformative projections and injections.

3 Modelling Paradigms and (Co)Algebra

The trajectories (time-indexed sets of contiguous states) of a dynamic system have been termed *recursive* by Rosen [15], but not in the rigorous sense of theoretical computer science. For discrete-time systems, where trajectories are sequences, the metaphor can be made precise by connecting finite/infinite trajectories with iteration/coiteration in the form of catamorphisms/anamorphisms, respectively. In this section, we discuss the transition from a philosophical view on modelling paradigms to formal systems that employ initial algebras and catamorphisms or final coalgebras and anamorphisms, respectively.

3.1 From Functional Modelling to Algebra

Our proposed mapping of the two modelling paradigms to (co)algebra is inspired by [15], where the functional paradigm is discussed in great philosophical detail and organized in the form of the commuting diagram depicted in Figure 2. We have adapted the original discussion to ecological problems in [8]. Note that the *real* side refers to the model system, not the target system. We identify the situation in this diagram with a pair of algebras, namely the real and the abstract one, with states as their elements, and the abstraction with a homomorphism. A model consists of

(i) the abstraction mapping that separates essential from accidental properties of real objects, and

(ii) a logical theory (system of equations) that specifies the valid progressions of abstract states.

The abstract algebra is merely a mathematical implementation of the specification. A scientific hypothesis is posed by claiming that the diagram commutes. Unlike in pure mathematics, and in the face of uncertainty about the model system, this is not a logical property to be decied, but rather an empirical property to be judged by testing and evidence, as prescribed by the Scientific Method. If the correspondence between the two sides actually holds, it gives clauses of the specification the special status of *laws of nature*. Reverting the top horizontal arrow results in the standard test situation for functional models, the *prediction*.

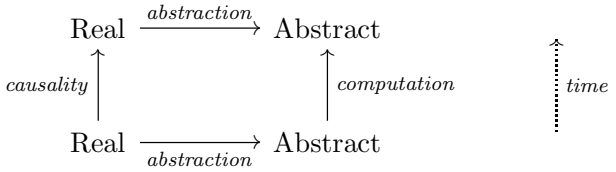


Fig. 2. Functional modelling, conceptually

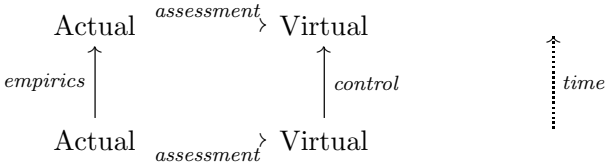


Fig. 3. Interactive modelling, conceptually

These philosophical interpretations need to be both formalized and generalized in order to adequately capture the tasks and capabilities of the functional modelling paradigm. We call algebraic modelling in the above, narow sense *direct* and distinguish it from *inverse* problems where not future states, but past or boundary conditions are investigated.

The basic tenet of algebraic modelling of both directions is to employ an initial algebra of a suitable functor as a formal *query language*, arbitrary algebras of the same functor (models) as implementations of query constructs and the catamorphisms as the recursive evaluation of queries.

3.2 From Interactive Modelling to Coalgebra

Changing the perspective from state to behavior affects all parts of the modelling situation. States are no longer required to be observable, but may be largely hidden behind an *interface*; all relevant information is taken from behavior at the interface. Metaphysically, objects and their properties are replaced by subjects and their actions. We reflect this shift of perspective, as common in the field of philosophy of science, by distinguishing the terms *real* (literally: of the things) and *actual* (literally: of the actions). Laws are replaced by their subjective counterparts, such as *strategies* and *norms*. Figure 3 shows the resulting commuting diagram. The standard test situation for interactive models is obtained by reversing the right vertical arrow; it describes *planning*. See Section 4.3 for a derivation of this model paradigm from formal representations.

The claim that coalgebraic modelling departs from the state-based perspective may be surprising. This issue arises from a fundamental difference between the notions of state in physics and in computer science. The observed state of a physical system *is* objective reality. The state of a formal automaton, as opposed to its physical implementation, merely *refers to* its actual behavior, in the sense that semantics are given in terms of observed transitions not states; for instance as the regular language accepted by a finite automaton. The reference character of state is expressed formally by the notion of bisimulation between alternative virtual systems

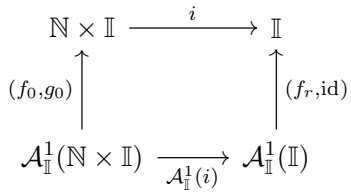


Fig. 4. Direct functional modelling (perfect information) with initial algebra

or by final coalgebraic semantics. We conjecture that this reflects the empirical phenomenon of *equifinality* [18,2]: The observed behavior of a complex system at a simpler interface can often be reconstructed by many different processes within the functional paradigm.

The basic tenet of coalgebraic modelling, in our sense, is to employ a final coalgebra as a formal *semantic domain*, arbitrary coalgebras of the same functor (models) as representatives of behavior and the anamorphisms as the recursive assessment of the represented behavior. The distinction between direct and inverse problems of coalgebraic modelling is less pronounced than in the algebraic case, at least for the example of the logistic map.

4 Formal Modelling Scenarios

4.1 Direct Functional Modelling

Direct functional modelling is a scenario where the "true" dynamics of a system are known. It solves the problem of *prediction*: From the observation of a current system state, future states are derived by formal (automatic) reasoning.

Claim 4.1 *The initial algebra of the functor $\mathcal{A}^1_B$, where B is the representation of system states, is paradigmatic for direct functional modelling.*

4.1.1 Perfect Information

The simplest case of direct functional modelling assumes perfect information about the precise current system state. Its application to the logistic map is shown in Figure 4. (Recall that the labels of vertical arrows are shorthands as defined in section 2.2.) The state space $\mathbb{I}$ is represented one-to-one. The left hand side is the simplified canonical initial $\mathcal{A}^1_{\mathbb{I}}$-algebra. The right hand side is a $\mathcal{A}^1_{\mathbb{I}}$-algebra that encodes the known dynamics of the system: Its carrier is the state space $\mathbb{I}$ and its operation is specified by the step function f_r (with $\mathrm{id}_{\mathbb{I}}$ as the trivial base case).

Theorem 4.2 *The catamorphism i for the operation specified by the pair $(f_r, \mathrm{id}_{\mathbb{I}})$ solves the problem of predicting a state n steps in the future, for $r \leq 4$.*

$$i(n, x) = f_r^n(x)$$

The preceding scenario is a straightforward reconstruction of the iterated step function f_r^n. The graph of the function consists of pairs of initial and final states, n steps apart; the intermediate states are forgotten. This can be remediated by

a simple refinement that replaces single states with stack-based representations of trajectories.

Definition 4.3 We define the set of **partial trajectories** as the set of stacks (strings constructed right-to-left) arising by iterated action of f_r on any initial state ($\mathcal{P}$ is the image functor).

$$T_r = \bigcup_{n=0}^{\infty} \mathcal{P}\big(\mathrm{push}(f_r)\big)^n(\mathbb{I}^1) \subseteq \mathbb{I}^+$$

- This is the smallest set such that $\mathbb{I}^1 \subseteq T_r$ and $\mathrm{push}(f_r) : T_r \to T_r$.

The refined model is shown in Figure 5. The carrier of the right hand side algebra is changed to T_r, and the operations f_r and $\mathrm{id}_{\mathbb{I}}$ have been replaced by $\mathrm{push}(f_r)$ and $\mathrm{in}_{\mathbb{I}}$, respectively, where $\mathrm{in}_A : A \to A^1$ is the injection of singleton strings.

Theorem 4.4 *The catamorphism i for the operation specified by the pair $\big(\mathrm{push}(f_r), \mathrm{in}_{\mathbb{I}}\big)$ solves the problem of predicting all states up to n steps in the future, for $r \leq 4$.*

$$i(n, x) = \mathrm{push}(f_r)^n(x)$$

The following three cases refine the representation of state and dynamics by replacing the state space $\mathbb{I}$ with progressively more complicated, derived spaces and replacing the step function f_r with an appropriate lifting to the respective space. Note that the requirement $r \leq 4$ is lifted.

4.1.2 Imperfect Information: Nondeterminism

A moderately simple case of direct functional modelling with imperfect information assumes nondeterminism. Note that the term "nondeterminism" is used in the usual sense of computer science, replacing the single precise current system state by a set of potential current system states. It is not used in the sense of philosophy, namely that a hidden variable, external source of randomness or decision-making entity is involved in the transition from one state to another.

The application of nondeterminstic direct functional modelling to the logistic map is shown in Figure 6. The state space $\mathbb{I}$ is represented by its powerset $\mathcal{P}(\mathbb{I})$. The left hand side is the simplified canonical initial $\mathcal{A}^1_{\mathcal{P}(\mathbb{I})}$-algebra. The right hand side is an $\mathcal{A}^1_{\mathcal{P}(\mathbb{I})}$-algebra that encodes the nondeterministic dynamics of the system: Its carrier is the set $\mathcal{P}(\mathbb{I})$ of sets of potential states and its operation is specified by

$$\begin{array}{ccc} \mathbb{N} \times \mathbb{I} & \overset{i}{\rightarrow} & T_r \\ {\scriptstyle (f_0, g_0)}\uparrow & & \uparrow{\scriptstyle (\mathrm{push}(f_r), \mathrm{in})} \\ \mathcal{A}^1_{\mathbb{I}}(\mathbb{N} \times \mathbb{I}) & \underset{\mathcal{A}^1_{\mathbb{I}}(i)}{\rightarrow} & \mathcal{A}^1_{\mathbb{I}}(T_r) \end{array}$$

Fig. 5. Direct functional modelling (partial trajectories) with initial algebra

$$\begin{array}{ccc} \mathbb{N} \times \mathcal{P}(\mathbb{I}) & \xrightarrow{\quad i \quad} & \mathcal{P}(\mathbb{I}) \\ {\scriptstyle (f_0, g_0)}\uparrow & & \uparrow{\scriptstyle \left(\mathcal{P}(f_r), \mathrm{id}\right)} \\ \mathcal{A}^1_{\mathcal{P}(\mathbb{I})}\left(\mathbb{N} \times \mathcal{P}(\mathbb{I})\right) & \xrightarrow[\mathcal{A}^1_{\mathcal{P}(\mathbb{I})}(i)]{} & \mathcal{A}^1_{\mathcal{P}(\mathbb{I})}\left(\mathcal{P}(\mathbb{I})\right) \end{array}$$

Fig. 6. Direct functional modelling (nondeterministic) with initial algebra

$\mathcal{P}(f_r)$, the image of state sets under f_r; a state is a potential post-state of a step if and only if it is the image of a potential pre-state under f_r.

Theorem 4.5 *The catamorphism i for the operation specified by the pair $\left(\mathcal{P}(f_r), \mathrm{id}_{\mathcal{P}(\mathbb{I})}\right)$ solves the problem of predicting a nondeterministic state n steps in the future.*

$$i(n, Y) = \mathcal{P}(f_r)^n(Y)$$

The nondeterministic case can be extended to more sophisticated imperfect information such as fuzzy sets of potential states.

4.1.3 Imperfect Information: Probabilism

Definition 4.6 Each continuous probability distribution over $\mathbb{I}$ is specified uniquely by a **cumulative distribution function (cdf)**, that is a continuous, weakly mononotic function $F : \mathbb{I} \to \mathbb{I}$ with $F(0) = 0$ and $F(1) = 1$.

- An $\mathbb{I}$-valued random variable X is said to be distributed according to F, written $X \sim F$, if and only if $F(y) = P(X \leq y) = P(X < y)$.
- We write $\widetilde{\mathbb{I}}$ for the set of cdfs over $\mathbb{I}$.

Definition 4.7 The function $\widetilde{f_r} : \widetilde{\mathbb{I}} \to \widetilde{\mathbb{I}}$ is defined as

$$\widetilde{f_r}(F)(y) = F\left(\tfrac{1}{2} - q_r(y)\right) + 1 - F\left(\tfrac{1}{2} + q_r(y)\right) \qquad q_r(y) = \begin{cases} \sqrt{\frac{1}{4} - \frac{y}{r}} & \text{if } y \leq \frac{r}{4} \\ 0 & \text{if } y > \frac{r}{4} \end{cases}$$

It is easy to verify that $\widetilde{f_r}(F)$ is in fact a cdf over $\mathbb{I}$. Note that $\frac{1}{2} \pm q_r(y)$ is the position of the vertical markers in Figure 1, right hand side.

Lemma 4.8 *The function $\widetilde{f_r}$ lifts a distribution over the function f_r.*

$$X \sim F \implies f_r(X) \sim \widetilde{f_r}(F)$$

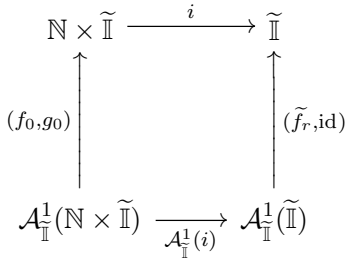


Fig. 7. Direct functional modelling (probabilistic) with initial algebra

Proof.

$$\begin{aligned}
P\big(f_r(X) \le y\big) &= P\big(rx(1-x) \le y\big) \\
&= P\Big(\big(x - \tfrac{1}{2}\big)^2 \ge \tfrac{1}{4} - \tfrac{y}{r}\Big) \\
&= P\Big(\big|x - \tfrac{1}{2}\big| \ge q_r(y)\Big) \\
&= P\big(x \le \tfrac{1}{2} - q_r(y) \vee x \ge \tfrac{1}{2} + q_r(y)\big) \\
&= P\big(x \le \tfrac{1}{2} - q_r(y)\big) + P\big(x \ge \tfrac{1}{2} + q_r(y)\big) \\
&= P\big(x \le \tfrac{1}{2} - q_r(y)\big) + 1 - P\big(x \le \tfrac{1}{2} + q_r(y)\big) \\
&= F\big(\tfrac{1}{2} - q_r(y)\big) + 1 - F\big(\tfrac{1}{2} + q_r(y)\big) \\
&= \widetilde{f}_r(F)(y)
\end{aligned}$$

□

The application of probabilistic direct functional modelling to the logistic map is shown in Figure 7. The state space $\mathbb{I}$ is represented by the set of cdfs $\widetilde{\mathbb{I}}$. The left hand side is the simplified canonical initial $\mathcal{A}^1_{\widetilde{\mathbb{I}}}$-algebra. The right hand side is a $\mathcal{A}^1_{\widetilde{\mathbb{I}}}$-algebra that encodes the probabilistic dynamics of the system: Its carrier is the set $\widetilde{\mathbb{I}}$ of state distributions and its operation is specified by $\widetilde{f}_r$, the action of f_r on the distribution of its argument.

Theorem 4.9 *The catamorphism i for the operation specified by the pair $(\widetilde{f}_r, \mathrm{id}_{\widetilde{\mathbb{I}}})$ solves the problem of predicting a probabilistic state n steps in the future.*

$$X \sim F \implies f_r^n(X) \sim i(n, F)$$

The probabilistic case can be extended to more complex, not purely continuous distributions.

4.2 Inverse Functional Modelling

Inverse functional modelling is a scenario where inferences about the dynamics of a system (parameters, initial or boundary conditions) are drawn from data recorded by external observation. It solves the problem of *reconstruction*: Empirical observations are reduced to possible causes (parameters and conditions not directly observable, but consistent with the data).

Claim 4.10 *The initial algebra of the functor $\mathcal{A}_1^A$, where A is the range of the observable system property of interest, is paradigmatic for inverse functional modelling.*

We choose the binary partition c as observable property. Its range is the binary alphabet 2, hence the carrier of the canonical initial algebra is the language of binary strings 2^*.

Definition 4.11 The function $w_r : \mathbb{N} \times \mathbb{I} \to 2^*$ is defined as

$$w_r(n, x) = c^*\big(\mathrm{push}(f_r)^{n-1}(x)\big)$$

It maps the pair (n, x) to the stack of observed binary symbols for n consecutive system states starting with x. Informally,

$$w_r(n, x) = c\big(f_r^{n-1}(x)\big) \cdots c\big(f_r^0(x)\big)$$

- The range of w_r for $n > 0$ is the set of partitioned partial trajectories $\mathcal{P}(c^*)(T_r)$.

The inverse modelling task, given data $w \in 2^*$ of length n and a parameter value r, is to find some or all solutions of the equation $w = w_r(n, x)$. A concise representation of the inferred information is given by a partial function on $\mathbb{I}$ that is defined only for initial states consistent with the observed data, and maps those to the final states after the observation. The solution is straightforwardly constructed, dealing with one observed symbol at a time.

Definition 4.12 We write $\overrightarrow{\mathbb{I}} = (\mathbb{I} \rightharpoonup \mathbb{I})$ for the space of partial functions on $\mathbb{I}$. The function $\overrightarrow{f_r} : 2 \times \overrightarrow{\mathbb{I}} \to \overrightarrow{\mathbb{I}}$ is defined as

$$\overrightarrow{f_r}(a, h) = f_r|_a \circ h \qquad \text{where} \quad f_r|_a = \begin{cases} f_r(x) & \text{if } c(x) = a \\ \text{undefined} & \text{if } c(x) \neq a \end{cases}$$

The operation $\overrightarrow{f_r}$ refines and extends a given partial function h by excluding initial states that are mapped by h to intermediate states inconsistent with a given data symbol a, and taking all others one f_r-step further.

The application of inverse functional modelling to the logistic map is shown in Figure 8. The observation range is the binary alphabet 2. The left hand side is the simplified canonical $\mathcal{A}_1^2$-algebra. The right hand side is a $\mathcal{A}_1^2$-algebra that encodes the elementwise refinement of inference: Its carrier, the "state space" of inference, is the set of partial functions $\overrightarrow{\mathbb{I}}$. Its operation is specified by $\overrightarrow{f_r}$, the action of f_r on the inference for its argument.

Theorem 4.13 *The catamorphism j for the operation specified by $(\overrightarrow{f_r}, \mathrm{id}_{\overrightarrow{\mathbb{I}}})$ solves the problem of inferring initial conditions from finite data.*

$$j(w)(x) = \begin{cases} f_r^n(x) & \textit{if } w = w_r(n, x) \\ \textit{undefined} & \textit{otherwise} \end{cases}$$

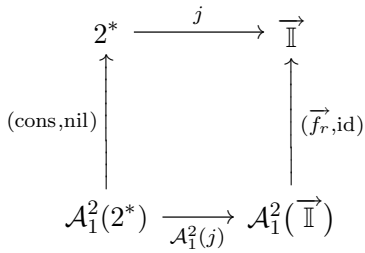


Fig. 8. Inverse functional modelling with initial algebra

Corollary 4.14 *The domain of $j(w)$ is a sound and monotonic approximate reconstruction of the initial state from finite data w, analogous to the method of nested intervals: Let $Y_r(n,x) = \mathrm{dom}\Big(j\big(w_r(n,x)\big)\Big)$. Then for all $m, n \geq 0$*

$$x \in Y_r(n,x) \qquad\qquad m < n \implies Y_r(m,x) \supseteq Y_r(n,x)$$

Here in the context of algebraic modelling, chronicles of events (observations of behavior) are used as means for identifying the initial and final state or the dynamics of the system under study. In the equation

$$j(a_1 \ldots a_n) = f_r|_{a_1} \circ \cdots \circ f_r|_{a_n}$$

however, an alternative view becomes apparent: The standard technique of category theory is to study objects without reference to their internal structure by studying the external structure of morphisms around them. Applied to the model above, this means studying the set of chronicles without reasoning about points of the transition functions they describe via j. This allows us to consider the limit $n \to \infty$, and represent complete, infinite behavior, for which the interpretation as end-to-end transition functions breaks down. This step takes us to the interactive paradigm on the philosophical level, and to coalgebra on the mathematical level.

4.3 Interactive Modelling

Interactive modelling is a scenario where the observable properties of a system are represented without referring to any particular process as their cause. It solves the problem of *assessment*: System states are no longer observed directly but classified according to their potential (future) behavior.

Claim 4.15 *The final coalgebra of the functor $\mathcal{A}_{\varnothing}^{A}$, where A is the range of the observable system property of interest, is paradigmatic for interactive modelling.*

For $r > 4$, the logistic map is not bounded by the interval $\mathbb{I}$; we treat the case that the interval is exceeded as undefined. The set $\mathrm{dom}(f_r)$ of points for which a single step is defined is easily characterized, but the set of points for which unboundedly many steps are defined is nontrivial. The following characterization and model are derived from [16].

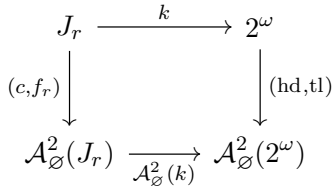


Fig. 9. Direct interactive modelling with final coalgebra

Definition 4.16 The set J_r such that $f_r^n(x) \in \mathbb{I}$ for all $x \in J_r$ and $n > 0$ is ($\overline{\mathcal{P}}$ is the preimage functor)

$$J_r = \bigcap_{n=0}^{\infty} \overline{\mathcal{P}}(f_r)^n(\mathbb{I}) \subseteq \mathbb{I}$$

- This is the largest set such that $J_r \subseteq \mathbb{I}$ and $f_r : J_r \to J_r$.
- If $r \leq 4$ then $J_r = \mathbb{I}$; otherwise J_r is a complicated (fractal) subset of $\mathbb{I}$.
- Note the duality to T_r in Definition 4.3.

Lemma 4.17 *For $r > 4$, the structure (J_r, γ) where γ is specified by (c, f_r) is a final $\mathcal{A}^2_\varnothing$-coalgebra.*

Proof. Section 18 of [16] gives an isomorphism $\tilde{c}$ between certain coalgebras over the category of complete metric spaces. Forgetting the metric structure, the following equations remain.

$$\begin{aligned} \mathrm{hd} \circ \tilde{c} &= c & \mathrm{tl} \circ \tilde{c} &= \tilde{c} \circ f_r \\ &= \mathrm{id}_2 \circ c \end{aligned}$$

Simple calculation yields

$$\begin{aligned} \underbrace{\iota_1 \circ \langle \mathrm{hd}, \mathrm{tl} \rangle}_{(\mathrm{hd},\mathrm{tl})} \circ \tilde{c} &= \big((\mathrm{id}_2 \times \tilde{c}) + \mathrm{id}_\varnothing\big) \circ \iota_1 \circ \langle c, f_r \rangle \\ &= \mathcal{A}^2_\varnothing(\tilde{c}) \circ \underbrace{\iota_1 \circ \langle c, f_r \rangle}_{(c,f_r)} \end{aligned}$$

That is, $\tilde{c}$ is a homomorphism, and hence isomorphism, between the coalgebras depicted in Figure 9. Since the right hand side is final, the left hand side is also final, and the isomorphism is the anamorphism k. □

Theorem 4.18 *The anamorphism k for the operation specified by the pair (c, f_r) solves the problem of representing the complete future behavior at the interface defined by c. Representations of the form $k(x)$ do not contain any reference to the parameter r or the initial value x.*

$$c\big(f_r^n(x)\big) = \mathrm{hd}\Big(\mathrm{tl}^n\big(k(x)\big)\Big) = k(x)(n)$$

This representation allows complete, infinite trajectories to be specified in the form $k^{-1}(s)$, in terms of a binary stream $s \in 2^\omega$. Empirical, finite data of behavior at

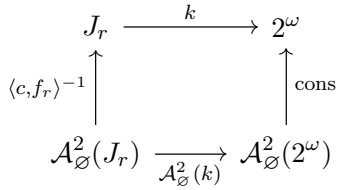


Fig. 10. Inverse interactive modelling with final coalgebra

the interface, formally collected using $w_r(n, x)$, is generally not sufficient to specify a trajectory uniquely in this way—an instance of the epistemological problem of induction; there is no logically safe procedure for obtaining nontrivial universal empirical truths [7,14]. This leads to a dual of the problem of measurement precision in state-based modelling, namely the problem of *complete* chronicles of behavior. A collection of data is complete in this sense if extrapolation from the observed strings to the possible streams is safe under given boundary conditions.

We have noted in Definition 1.2 that the operation $\langle c, f_r \rangle$ is bijective on both $\mathbb{I}$ and the subset J_r for $r \geq 4$. Incidentally, the latter is the operation of the $\mathcal{A}^2_1$-coalgebra depicted on the left hand side of Figure 9. Since the operation of the final $\mathcal{A}^2_1$-coalgebra is also bijective (by Lambek's Lemma), we may reverse the vertical arrows to arrive at the diagram shown in Figure 10. Note that the distinction between algebra and coalgebra is rather blurred in this scenario.

The operation $\langle c, f_r \rangle^{-1} : 2 \times J_r \to J_r$ models a non-autonomous dynamic system with binary input in each step. This input may be interpreted as the nondeterministic *choice* of an agent, either internal or external to the system. Under this interpretation, prediction is no longer a valid problem. But this apparent restriction is actually a trade-off: On the upside, it becomes possible to investigate actually observed, contingent behavior in virtualized form in terms of subcoalgebras of the final coalgebra. Laws regarding the presence or absence of certain patterns in these subsystems, described by a theory in modal logic, reflect *strategies* in the actual system, the dual of natural laws. Examples of the relevance of strategies, both literally and figuratively, abound in ecology: Organisms prefer favourable and avoid hostile environments, natural selection is most effectively described in strategic terms, ecosystem use is governed by economic rationale and social norms; cf. the domestication example in section 1.1.

5 Conclusion

We have demonstrated that for the simple logistic model, the relationship between the functional and the interactive modelling paradigm can be made formally precise as the algebra–coalgebra dualism. Since dualism is not equivalence, the key issue for further research is where the two approaches deviate, both on the empirical level regarding the role of data and on the theoretical level regarding the role of formalisms. The keywords of both paradigms are given in Table 1 in synopsis.

In ecology and environmental sciences, the functional paradigm is prevalent but not unconditionally successful [13]. Therefore, the added value of interactive models

Table 1
Modelling paradigms and keywords

Paradigm	Functional	Interactive
Ontological Basis	state	behavior
Origin of Formalisms/Metaphors	physics	computer science
Empirical Reference	simple building blocks w. invariants	complete behavior history w. utilization record
Tests	prediction reconstruction	assessment planning
Mathematic Structure	algebra	coalgebra
Logic	equational	modal
Theory Example	energy conservation	sustainable use
Application Domains	geosciences weather forecast	simulation, games ecosystem management

is of particular interest. Many essential features of living systems are naturally characterized in terms of behavior, e.g. feeding, reproducing, growing, evolving. Being alive is not a state property in the functional sense, as the development towards artificial life has shown [3]. Coalgebraic modelling facilitates the formal organization of chronicles, as opposed to measurements; this may prove an important extension in this context.

Interactive theories formulated in coalgebra not only have a different formal presentation, they encode different pragmatics. On the functional side, problems of prediction and reconstruction are solved by searching for laws that govern the dynamic mechanism. On the interactive side, problems of assessment and management are solved by searching for strategies, norms or intentions that govern the behavior of agents. The transition from the former to the latter paradigm will not solve the notoriously difficult problems about explaining ecosystems, but offers the opportunity to formalize models of sustaining ecosystems.

5.1 Related Work

The inspiration to use the logistic map to demonstrate the potential of the algebra–coalgebra duality for scientific modelling has been taken from [16], where the result that forms the foundation of our interactive modelling scenario is given rather in passing.

The characterization of functional and interactive modelling as commutative diagrams has been given in [17], where we have criticised the situation of theoretical ecology from the perspective of software science.

The technique of realizing (co)recursive operations as cata-/anamorphisms of simpler operations has been adapted from the *Squiggol* approach to constructive functional programming; confer the famous banana notation of [11].

References

[1] Berryman, A., *The origins and evolution of predator–prey theory*, Ecology **73** (1992), pp. 1530–1535.

[2] Beven, K. J., *A manifesto for the equifinality thesis*, Journal of Hydrology **320** (2006), pp. 18–36.

[3] Brooks, R., *The relationship between matter and life*, Nature **409** (2001), pp. 409–411.

[4] Crutchfield, J. P., *Observing complexity and the complexity of observation*, in: H. A. Atmanspacher and G. J. Dalenoort, editors, *Inside versus Outside*, Springer Series in Synergetics, Springer-Verlag, Berlin, 1994 pp. 235–272.

[5] Ehrig, H. and B. Mahr, "Fundamentals of Algebraic Specification I. Equations and Initial Semantics," Springer-Verlag, Berlin, 1985.

[6] Frigg, R., *Fiction and scientific representation*, in: R. Frigg and M. Hunter, editors, *Beyond Mimesis and Nominalism: Representation in Art and Science*, Springer-Verlag, Berlin, 2009 .

[7] Goodman, N., "Fact, Fiction, & Forecast," Harvard University Press, 1955.

[8] Hauhs, M. and H. Lange, *Foundations for the simulation of ecosystems*, in: J. Lenhard, G. Küppers and T. Shinn, editors, *Simulation: Pragmatic Constructions of Reality*, number 25 in Sociology of the Sciences Yearbook, Kluwer Academic Publishers, Dordrecht, 2006 pp. 57–77.

[9] Jacobs, B., *Exercises in coalgebraic specification*, in: *Algebraic and coalgebraic methods in the mathematics of program construction*, Springer-Verlag, New York, 2002 pp. 237–280.

[10] May, R. M., *Simple mathematical models with very complicated dynamics*, Nature **261** (1976), pp. 459–467.

[11] Meijer, E., M. Fokkinga and R. Paterson, *Functional programming with bananas, lenses, envelopes and barbed wire*, in: *Proceedings of the 5th ACM conference on Functional programming languages and computer architecture* (1991), pp. 124–144.

[12] Murray, J. D., "Mathematical Biology," Springer-Verlag, Heidelberg, 1989.

[13] Peters, R. H., "A Critique for Ecology," Cambridge University Press, 1991.

[14] Popper, K. R. and D. W. Miller, *A proof of the impossibility of inductive probability*, Nature **302** (1983), pp. 687–688.

[15] Rosen, R., "Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life," Columbia University Press, New York, 1991.

[16] Rutten, J. J. M. M., *Universal coalgebra: a theory of systems*, Theor. Comput. Sci. **249** (2000), pp. 3–80.

[17] Trancón y Widemann, B. and M. Hauhs, *Programming as a model for the theory of ecosystems*, in: J. Knoop and A. Prantl, editors, *Post-Proceedings of 15. Kolloquium Programmiersprachen und Grundlagen der Programmierung*, Technische Universität Wien, 2009, to appear.

[18] von Bertalanffy, L., "General Systems Theory: Foundations, Development, Applications," 1968.